

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein-search, using sw model

Run on: March 5, 2002, 17:05:02 ; Search time 23.46 Seconds
(without alignments)
918.812 Million cell updates/sec

Title: US-09-424-705-2
Perfect score: 1536
Sequence: 1 MKYLPTAAAGLLILAAQPA.....GSEQKILSEEDLNHHNNHH 291

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

A_Geneseq_1101.*
1: /SID88/gcgdata/geneseq/geneseqp/AA1980.DAT.*
2: /SID88/gcgdata/geneseq/geneseqp/AA1981.DAT.*
3: /SID88/gcgdata/geneseq/geneseqp/AA1982.DAT.*
4: /SID88/gcgdata/geneseq/geneseqp/AA1983.DAT.*
5: /SID88/gcgdata/geneseq/geneseqp/AA1984.DAT.*
6: /SID88/gcgdata/geneseq/geneseqp/AA1985.DAT.*
7: /SID88/gcgdata/geneseq/geneseqp/AA1986.DAT.*
8: /SID88/gcgdata/geneseq/geneseqp/AA1987.DAT.*
9: /SID88/gcgdata/geneseq/geneseqp/AA1988.DAT.*
10: /SID88/gcgdata/geneseq/geneseqp/AA1989.DAT.*
11: /SID88/gcgdata/geneseq/geneseqp/AA1990.DAT.*
12: /SID88/gcgdata/geneseq/geneseqp/AA1991.DAT.*
13: /SID88/gcgdata/geneseq/geneseqp/AA1992.DAT.*
14: /SID88/gcgdata/geneseq/geneseqp/AA1993.DAT.*
15: /SID88/gcgdata/geneseq/geneseqp/AA1994.DAT.*
16: /SID88/gcgdata/geneseq/geneseqp/AA1995.DAT.*
17: /SID88/gcgdata/geneseq/geneseqp/AA1996.DAT.*
18: /SID88/gcgdata/geneseq/geneseqp/AA1997.DAT.*
19: /SID88/gcgdata/geneseq/geneseqp/AA1998.DAT.*
20: /SID88/gcgdata/geneseq/geneseqp/AA1999.DAT.*
21: /SID88/gcgdata/geneseq/geneseqp/AA2000.DAT.*
22: /SID88/gcgdata/geneseq/geneseqp/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1354	88.2	539	21	AAV50823
2	1346.5	87.7	554	21	AAV50822
3	1300.5	84.7	288	19	AAW82482
4	1208.5	78.7	288	19	AAW82316
5	1208.5	78.7	288	19	AAW82317
6	1132	73.7	531	21	AAV43749
7	1010.5	65.8	281	20	AAW82744
8	999.5	65.1	330	22	AAW80842
9	974	63.4	246	15	AAW60522
10	964.5	62.8	309	22	AAW70841
11	963.5	62.7	239	21	AAW12563

12	963.5	62.7	239	22	AAW36826	Anti-TAC sfv prote
13	957.5	62.3	599	19	AAW37132	Anti-Tac(Fv)-PE40
14	957.5	62.3	599	21	AAW87477	Anti-Tac(Fv)-PE40
15	952.5	62.0	599	20	AAW95462	Anti-Tac(Fv)-PE40
16	947	61.7	288	20	AAW82743	Fusion protein PNG
17	940	61.2	291	22	AAW20443	Antibody 8860-myc
18	928.5	60.4	270	16	AAW75719	MFE-23 antibody
19	919	59.8	244	16	AAW79867	Anti-EGFR single c
20	918	59.8	305	21	AAV59265	Antibody 4H5 L cha
21	918	59.8	305	21	AAV51142	Murine derived pro
22	916	59.6	242	16	AAW79872	Anti-EGFR single c
23	914.5	59.5	423	22	AAW30695	A fusion of anti-C
24	912	59.4	267	20	AAW85688	Humanised D9D10 sc
25	911	59.3	244	16	AAW79873	Anti-EGFR single c
26	908	59.1	246	16	AAW79869	Anti-EGFR single c
27	905	58.9	244	16	AAW79868	Anti-EGFR single c
28	904	58.9	241	22	AAW50019	Antimesothelin ant
29	901.5	58.7	483	19	AAW82315	Chimeric CD19/crCR
30	901	58.7	242	16	AAW79870	Anti-EGFR single c
31	898	58.5	553	18	AAW11508	Single chain anti-
32	898	58.5	553	22	AAW73223	H22-anti-CEA antib
33	898	58.5	553	22	AAW85455	Bispecific single
34	898	58.5	553	22	AAW61960	Bispecific single
35	896.5	58.4	322	22	AAW20440	Antibody 8860 Diva
36	895.5	58.3	281	19	AAW82314	Human ScFvCD19 pro
37	894	58.2	276	17	AAW02227	Anti-B7.1/anti-B7.
38	889.5	57.9	271	17	AAW02293	FKC-II-KEX2. Chl
39	888.5	57.8	242	19	AAW53881	Firefly luciferase
40	886	57.7	242	16	AAW79871	Anti-EGFR single c
41	886	57.7	263	16	AAW90226	Anti-B7.2 monospec
42	883.5	57.5	262	22	AAW70762	Single chain Fv an
43	883.5	57.5	268	20	AAW90222	Anti-B7.2 monospec
44	883.5	57.5	281	20	AAW90223	Anti-B7.1/anti-B7.
45	881.5	57.4	264	18	AAW35562	HindIII-EcoRI inse

ALIGNMENTS

RESULT 1	
ID	AAV50823 standard; Protein; 539 AA.
XX	
AC	AAV50823;
XX	
DF	18-FEB-2000 (first entry)
XX	
DE	Fv-antibody construct containing antibody 9E10 epitope protein.
XX	
KW	Multivalent; Fv-antibody; antiviral; antibacterial; cytostatic;
KW	diagnosis; therapy; disease.
XX	
OS	Synthetic.
XX	
PN	DE19819846-A1.
XX	
PD	11-NOV-1999.
XX	
PF	05-MAY-1998; 98DE-1019846.
XX	
PR	05-MAY-1998; 98DE-1019846.
XX	
PA	(DEKR-Y) DEUT KREBSFORSCHUNGSZENTRUM.
XX	
PI	Little M, Kipriyanov S;
XX	
DR	WPI: 2000-024472/03.
XX	
XX	N-PSDB; AA243432.
PT	Multivalent Fv-antibody constructs with at least four variable domains
PT	connected by 1, 2 and 3 peptide linkers
XX	
PS	Example 1; Fig 6; 14pp; German.

XX This invention describes a novel multivalent Fv-antibody construct with
 CC at least four variable domains that are connected to one another by 1,
 CC 2 and 3 peptide linkers. The construct has antiviral, antibacterial
 CC and cytostatic activity. The multivalent Fv-antibody constructs are
 CC useful for the diagnosis and/or therapy of disease, especially viral,
 CC bacterial or tumor diseases. The multivalent Fv-antibody constructs have
 CC increased stability when in the form of a single chain dimer. This
 CC sequence represents a tetraivalent Fv antibody construct composed of the
 CC antibody 9E10 epitope in expression plasmid pDISC3x19-SL

XX Sequence 539 AA;

Query Match 88.2%; Score 1354; DB 21; Length 539;

Best Local Similarity 52.9%; Pred. No. 3.5e-80; Mismatches 2; Indels 248; Gaps 3;

Matches 285; Conservative 2; Mismatches 4; Indels 248; Gaps 3;

1 MKYLLPTAAAGLLLLAOPAMAQVOLOQSGAELARPGASVKMSCKASGTTFRITMHWK 60
 1 mkyllptaaglllllaaqpamaqvqlqsgaelarpgasvkmckasgytfritylmhwk 60

61 QRPQGLEWIGYINPSRGYTNQKFKDKATLTTPDKSSSTAYMOUSLTSEDSAVYYCAR 120
 61 qrpqglewiyinpsrgytnyqkfkdaklttdkssstaymqsltsedsavyycar 120

61 QRPQGLEWIGYINPSRGYTNQKFKDKATLTTPDKSSSTAYMOUSLTSEDSAVYYCAR 120
 61 qrpqglewiyinpsrgytnyqkfkdaklttdkssstaymqsltsedsavyycar 120

121 YYDDHYSIDYWGQGTFLVSSAKTTPKL----- 148
 121 yyddhysidywqgtflvssakttpklvgdilltqtpslavslgqratlsckasqsvd 180

121 yyddhysidywqgtflvssakttpklvgdilltqtpslavslgqratlsckasqsvd 180

149 ----- 148

181 ydgdsylmwyqqlpqgpklllydesnlvsjiprfsgsgtdftlnhpkykdaaly 240

149 ----- 148

241 hcqgstedpwtftggtkleikradaaaagpgsgvqlqsgaelvrgpsvskiskasy 300

149 ----- 153

301 afssymmwkqrpqglewlgqlwpgdgdlnyngkfkykaltladesstaymqslsla 360

154 ----- 172

361 sedasavfcarrettlvgryyamdywgqstsvtsasakttpkllvgdilltqspalmes 420

154 ----- 172

SEARV-----DIVLTQSPAIMAS 172

173 PGEKVTMTCSASSSYTNMWTQOKSGTSPKRMITDTSKLASGVPAHFRSGSGTSTSLTI 232
 173 pgekvmtcsasssytnmwtqoksgtspkrmitdtsklasgvpahfrsgsgtstslti 232

421 pgekvtmtcsasssytnmwtqoksgtspkrmitdtsklasgvpahfrsgsgtstslti 480

233 SGMEADATYCCOOWSSNPFEGSGTKLEINRADTAPRGSEOKLTISEEDLNSHHHHH 291
 233 sgmeadaatyccowssnpfegsgtkleinradtaprgseokltiseedlnshhhhh 291

481 sgmeadaatyccowssnpfegsgtkleinradtaprgseokltiseedlnshhhhh 539

RESULT 2
 ID AAY50822 standard; Protein: 554 AA.
 AC AAY50822;
 XX 18-FEB-2000 (first entry)
 DE Fv-antibody construct containing antibody 9E10 epitope protein.
 KW Multivalent; Fv-antibody; antiviral; antibacterial; cytostatic;
 KW diagnosis; therapy; disease.
 OS Synthetic.
 XX DEL19819846-A1.
 PD 11-NOV-1999.

XX 05-MAY-1998; 98DE-1019846.
 PF 05-MAY-1998; 98DE-1019846.
 PR (DEKR-) DEUT KREBSFORSCHUNGSZENTRUM.
 PA Little M, Kipriyanov S;
 PI WPI: 2000-024472/03.
 DR N-PSDB: AA243431.
 DR Multiivalent Fv-antibody constructs with at least four variable domains
 CC connected by 1, 2 and 3 peptide linkers
 CC Example 1; Fig 5; German.

XX This invention describes a novel multivalent Fv-antibody construct with
 CC at least four variable domains that are connected to one another by 1,
 CC 2 and 3 peptide linkers. The construct has antiviral, antibacterial
 CC and cytostatic activity. The multivalent Fv-antibody constructs are
 CC useful for the diagnosis and/or therapy of disease, especially viral,
 CC bacterial or tumor diseases. The multivalent Fv-antibody constructs have
 CC increased stability when in the form of a single chain dimer. This
 CC sequence represents a bivalent Fv antibody construct composed of the
 CC antibody 9E10 epitope in expression plasmid pDISC3x19-TL.

XX Sequence 554 AA;

Query Match 87.7%; Score 1346.5; DB 21; Length 554;
 Best Local Similarity 51.4%; Pred. No. 1.1e-79;
 Matches 285; Conservative 2; Mismatches 4; Indels 263; Gaps 3;

1 MKYLLPTAAAGLLLLAOPAMAQVOLOQSGAELARPGASVKMSCKASGTTFRITMHWK 60
 1 mkyllptaaglllllaaqpamaqvqlqsgaelarpgasvkmckasgytfritylmhwk 60

61 QRPQGLEWIGYINPSRGYTNQKFKDKATLTTPDKSSSTAYMOUSLTSEDSAVYYCAR 120
 61 qrpqglewiyinpsrgytnyqkfkdaklttdkssstaymqsltsedsavyycar 120

61 QRPQGLEWIGYINPSRGYTNQKFKDKATLTTPDKSSSTAYMOUSLTSEDSAVYYCAR 120
 61 qrpqglewiyinpsrgytnyqkfkdaklttdkssstaymqsltsedsavyycar 120

121 YYDDHYSIDYWGQGTFLVSSAKTTPKL----- 148
 121 yyddhysidywqgtflvssakttpklvgdilltqtpslavslgqratlsckasqsvd 180

121 yyddhysidywqgtflvssakttpklvgdilltqtpslavslgqratlsckasqsvd 180

149 ----- 148

181 ydgdsylmwyqqlpqgpklllydesnlvsjiprfsgsgtdftlnhpkykdaaly 240

149 ----- 148

241 hcqgstedpwtftggtkleikradaaaagpgsgvqlqsgaelvrgpsvskiskasy 300

149 ----- 153

301 afssymmwkqrpqglewlgqlwpgdgdlnyngkfkykaltladesstaymqslsla 360

154 ----- 158

SEARV-----DIVLTQSPAIMAS 158

159 -DIVLTQSPAIMASAPGKVTMTCSASSSYTNMWTQOKSGTSPKRMITDTSKLASGVA 217
 159 -divltqspaimasapgekvtmtcsasssytnmwtqoksgtspkrmitdtsklasgva 217

421 gdlvltqspaimasapgekvtmtcsasssytnmwtqoksgtspkrmitdtsklasgva 480

218 HFRSGSGTSTSLTIISGMEADATYCCOOWSSNPFEGSGTKLEINRADTAPRGSEOKLT 277
 218 hfrsgsgtstsltiisgmeadaatyccowssnpfegsgtkleinradtaprgseoklt 277

481 hfrsgsgtstsltiisgmeadaatyccowssnpfegsgtkleinradtaprgseoklt 540

278 ISEEDLNSHHHHH 291
 278 iseedlnshhhhh 554

```
RESULT 3
AAW82482
ID AAW82482 standard; Protein: 288 AA.
XX
AC AAW82482:
XX
DT 26-FEB-1999 (first entry)
XX
DE Mouse bispecific antibody variant OKT3/anti-CD3 protein.
XX
KW OKT3; monoclonal antibody; MAb; point mutation; transplant rejection;
KW organ recipient; diagnosis; tumour; therapy; antibody; anti-CD19;
KW anti-CD3.
XX
OS Mus sp.
OS Synthetic.
XX
PN DE19721700-C1.
XX
PF 19-NOV-1998.
XX
PR 23-MAY-1997; 97DE-1021700.
XX
PR 23-MAY-1997; 97DE-1021700.
XX
PA (DEKR-) DEUT KREBSFORSCHUNGSZENTRUM.
XX
PI Kipriyanov S, Little M, Moldenhauer G;
XX
DR WPI; 1998-596150/51.
XX
PT Monoclonal antibody OKT3 with point-mutation - where cysteine is
PT replaced by another polar amino acid, useful for controlling
PT transplant rejection, and in tumour diagnostics and therapy
XX
PS Disclosure; Fig 3; 8pp; German.
XX
CC This sequence represents a protein which has anti-CD3 activity and is
CC encoded by a monoclonal antibody (MAb) diabody derived from OKT3 with a
CC point-mutation where Cys at position H100A is replaced with another polar
CC amino acid, in this example Ser. The diabody encodes two OKT3 proteins,
CC one which has anti-CD19 activity and one with anti-CD3 activity. The MAb
CC is used in lowering or eliminating the transplant rejection in an organ
CC recipient and for diagnostic methods for tumours and tumour therapy.
XX
SQ Sequence 288 AA:
SO
Query Match 84.7%; Score 1300.5; DB 19; Length 288;
Best Local Similarity 85.3%; Pred. No. 5.2e-77;
Matches 255; Conservative 10; Mismatches 15; Indels 19; Gaps 3;
QY 1 MKYLLPTAAGLLLLAOPAMAQVOLOOSGAELARPGASVKNKSCASGYTFRITMHWK 60
DB 1 mkyllptaaglllllaopamaqvqlqsgaelvrgpsvskisckasyafesymmwk 60
QY 61 ORPGGLEWIGVINSRGVTNNOKFKDKATITTPKSSSTAYMOLSLTSEDSAYYCAR 119
DB 61 qrpqglwvlgwipgddltynqgfkqkaltadessstaymqslsasedsavyfcar 120
QY 120 -----RYDDHYSLDYGQGTTLTVSSAKTTPKLEEGEFSEARVDIVLTQSPAIMSAS 172
DB 121 rettvtgryy---yamdywgqgtsvltvasaktlprk19g-----dlvtlqspaimsas 169
QY 173 PGEKVTMTCSASSSVSYNMWYQOKSGTSPKRWIYDTSKLASGVPAHFRGSGSGTSYSLTI 232
DB 170 pgekvmtcasssvsyymwvyqgkqetaprkrcwlydtsklaagvphftfsgsgtetsylti 229
QY 233 SCMEDEADATYYCOOWSSNPTFGSGTKLEINRADTAPGSGOKLISEEDLNSHHHNNH 291
DB 230 sgmeadeadaatyccqgwsnptfsgtklelnradtaptgseqkliseedlnshhhnh 288
```

```
RESULT 4
AAW82316
ID AAW82316 standard; Protein: 288 AA.
XX
AC AAW82316:
XX
DT 26-FEB-1999 (first entry)
XX
DE Mouse OKT3 variant antibody protein.
XX
KW OKT3; monoclonal antibody; MAb; point mutation; transplant rejection;
KW organ recipient; diagnosis; tumour; therapy.
XX
OS Mus sp.
OS Synthetic.
XX
PN DE19721700-C1.
XX
PF 19-NOV-1998.
XX
PR 23-MAY-1997; 97DE-1021700.
XX
PR 23-MAY-1997; 97DE-1021700.
XX
PA (DEKR-) DEUT KREBSFORSCHUNGSZENTRUM.
XX
PI Kipriyanov S, Little M, Moldenhauer G;
XX
DR WPI; 1998-596150/51.
DR N-PSDB; AAV73335.
XX
PT Monoclonal antibody OKT3 with point-mutation - where cysteine is
PT replaced by another polar amino acid, useful for controlling
PT transplant rejection, and in tumour diagnostics and therapy
XX
PS Claim 3; Fig 2; 8pp; German.
XX
CC This sequence represents a monoclonal antibody (MAb) derived from OKT3
CC with a point-mutation where Cys at position H100A is replaced with
CC another polar amino acid, in this example Ser. The MAb is used in
CC lowering or eliminating the transplant rejection in an organ recipient,
CC and for diagnostic methods for tumours and tumour therapy.
XX
SQ Sequence 288 AA:
SO
Query Match 78.7%; Score 1208.5; DB 19; Length 288;
Best Local Similarity 78.7%; Pred. No. 4.7e-71;
Matches 233; Conservative 17; Mismatches 33; Indels 13; Gaps 2;
QY 1 MKYLLPTAAGLLLLAOPAMAQVOLOOSGAELARPGASVKNKSCASGYTFRITMHWK 60
DB 1 mkyllptaaglllllaopamaqvqlqsgaelarpgasvknkscasytlfrylmhwk 60
QY 61 ORPGGLEWIGVINSRGVTNNOKFKDKATITTPKSSSTAYMOLSLTSEDSAYYCAR 120
DB 61 qrpqglwvlgwipgddltynqgfkqkaltadessstaymqslsasedsavyfcar 120
QY 121 YDDHYSLDYGQGTTLTVSSAKTTPKLEEGEFSEARVDIVLTQSPAIMSASPGKVTMT 180
DB 121 yddhyssldywgqgtllvssaktlprk19g-----dliltqspaimslavslgqratis 172
QY 181 CSASSV-----SYNMWYQOKSGTSPKRWIYDTSKLASGVPAHFRGSGSGTSYSLTIGM 235
DB 173 cksasvsvydgdsylnwvqilp9gppklliydasnlvsgipprfifsgsgsgdftlnlhpv 232
QY 236 EAEDEADATYYCOOWSSNPTFGSGTKLEINRADTAPGSGOKLISEEDLNSHHHNNH 291
DB 233 ekvdaatyhcqgstedpwtlfggktleikrdaaagaegklliseedlnshhhnh 288
```

RESULT 5
 AAM82317 standard; Protein: 288 AA.
 XX
 AC AAM82317;
 XX
 DT 26-FEB-1999 (first entry)
 XX
 DE Mouse bispecific antibody variant OKT3/anti-CD19 protein.
 XX
 KW OKT3; monoclonal antibody; Mab; point mutation; transplant rejection;
 KW organ recipient; diagnosis; tumour; therapy; diabody; anti-CD19;
 KM anti-CD3.
 XX
 OS Mus sp.
 OS Synthetic.
 XX
 PN DE19721700-CL.
 19-NOV-1998.

PE 23-MAY-1997; 97DE-1021700.
 XX
 PR 23-MAY-1997; 97DE-1021700.
 XX
 PA (DEKR-) DEUT KREBSFORSCHUNGSZENTRUM.
 PI Kipriyanov S, Little M, Moldenhauer G;
 XX
 DR WPI; 1998-596150/51.
 DR N-PSDB; AAV73337.
 XX
 PT Monoclonal antibody OKT3 with point-mutation - where cysteine is
 PT replaced by another polar amino acid, useful for controlling
 PT transplant rejection, and in tumour diagnostics and therapy
 PS
 PS Disclosure; Fig 3; 8pp; German.

This sequence represents a protein which has anti-CD19 activity and is encoded by a monoclonal antibody (Mab) diabody derived from OKT3 with a point-mutation where Cys at position H100A is replaced with another polar amino acid, in this example Ser. The diabody encodes two OKT3 proteins, one which has anti-CD19 activity and one with anti-CD3 activity. The Mab is used in lowering or eliminating the transplant rejection in an organ recipient and for diagnostic methods for tumours and tumour therapy.

Sequence 288 AA:

Query Match 78.7%; Score 1208.5; DB 19; Length 288;
 Best Local Similarity 78.7%; Pred. No. 4.7e-71;
 Matches 233; Conservative 17; Mismatches 33; Indels 13; Gaps 2;

QY 1 MKYLLPTAAAGILLAAQPAQAQVQLQSGAEELARPAGASVKNKSCASGYTFRYTMHWK 60
 DB 1 mkyllptaagilllaaqpamaqvqlqsgaeelarpagsvknscasgytfrlytmhwk 60
 QY 61 QRPQGLEWIGYINPSRGITNPNKFKDKATLTTRKSSSTAWQLSSLTSESAVYYCAR 120
 DB 61 qrpqglewiyinpsrgitnpgkfkdaltltkssstaymqslsseasavyear 120
 QY 121 YYDHYSLDYWGQFTLLVSSAKTPKLEEGEFSEARVDIVTOSPAIMASPGEKVMT 180
 DB 121 yydhyssldywqgftllvssaktprkleggefsearvdivtospaimaspgekvmt 180
 QY 121 YYDHYSLDYWGQFTLLVSSAKTPKLEEGEFSEARVDIVTOSPAIMASPGEKVMT 172
 DB 121 yydhyssldywqgftllvssaktprkleggefsearvdivtospaimaspgekvmt 172
 QY 181 CSASSSV-----SYNMWYQOKSGTSPKRWIYDTSLASGVPAHFPGSGSGTSYSLTISGM 235
 DB 173 ckaasgvydgdsgylnwqyqpgpkllydaanlvsqiprfsgsgstftlnlhpv 232
 QY 236 EAEDAAITVCCQWSSNPTFGSGTRKLEINRADTAPGSEOKLISEDLNHHHHH 291
 DB 233 ekvaatvchcgsqstedpwtfgsgtklekrdaaaseqkliseedlnshhhhh 288

RESULT 6
 AAY43749 standard; Protein: 531 AA.
 XX
 AC AAY43749;
 XX
 DT 11-FEB-2000 (first entry)
 XX
 DE Amino acid sequence of the bscCD19xCD3 antibody.
 XX
 KW bscCD19xCD3 antibody; bispecific single-chain fragment; CD19 antigen;
 KW CD3 antigen; CD19-positive target cell; T-cell stimulation;
 KW cytotoxic T-lymphocyte; B-cell malignancy; myasthenia gravis;
 KW B-cell mediated autoimmune disease; Morbus Basedow;
 KW Hashimoto thyroiditis; Goodpasture syndrome; B-cell depletion;
 KW non-Hodgkin lymphoma; gene therapy; cancer; viral disease.
 XX
 OS Synthetic.
 XX
 FT Key location/Qualifiers
 FH
 FT Peptide 1..19
 FT /note= "signal peptide"
 FT 28..138
 FT /note= "VL-CD19 peptide"
 FT 154..277
 FT /note= "VH-CD19 peptide"
 FT 283..401
 FT /note= "VH-CD3 peptide"
 FT 420..525
 FT /note= "VL-CD3 peptide"
 FT 526..531
 FT /note= "his tag"

W09954440-A1.
 28-OCT-1999.
 21-APR-1999; 99WO-EP02693.
 21-APR-1998; 98EP-0107269.
 (DOER/) DOERKEN B.
 (RIET/) RIETHUELLER G.
 Kufer P, Lutterbuese R, Bargou R, Loeffler A;
 WPI; 2000-013241/01.
 N-PSDB; AAZ30332.
 Novel multifunctional polypeptide for treating B-cell malignancies especially non-Hodgkin lymphoma

Example 2; Fig 8; 91pp; English.

The present sequence represents a bscCD19xCD3 antibody. This antibody is a bispecific single-chain polypeptide comprising domains providing binding-site of immunoglobulin chains or antibodies specifically recognizing CD19 and CD3 antigen. The polypeptide destroys CD19-positive target cells without any need of T-cell pre and/or co-stimulation, by recruiting cytotoxic T-lymphocytes and so specific lysis by T-cells rather than a direct effect by an antibody is achieved. The bispecific single-chain polypeptides, or nucleotides encoding them, are used for the treatment of B-cell malignancies, B-cell mediated autoimmune diseases like myasthenia gravis, Morbus Basedow, Hashimoto thyroiditis or Goodpasture syndrome or for the depletion of B-cells and more particularly non-Hodgkin lymphoma in mammals preferably human. They can also delay the pathological conditions caused by these diseases, and can be used for detecting these diseases. The polynucleotide is used for gene therapy. The polypeptides are also used for identifying compounds modulating B-cell/T-cell mediated immune response which can in turn be used for treating cancer, its related diseases and also for inhibiting viral diseases by preventing viral infection.

XX	SO	Sequence	531 AA;
XX	Query Match	73.7%; Score 1132; DB 21; Length 531;	
XX	Best Local Similarity	80.1%; Pred. No. 8.1e-66;	
XX	Matches 217; Conservative	7; Mismatches 25; Indels 22; Gaps	2
YY	22 AOVLOOSGAEIARPGASVKKMSCKASGTFETRYTMMVKQRPQGLEWITGIYNPSRGYTN	81	
YY	82 YNOEFKDKATLTPTKSSSTAAVMQLSLTSEDSAAVYGCARYYDDHSLDVGSGTTLTVSS	141	
YY	342 yndqfkdadallctdksstcaymqdsslsedsaayycaryyddhycldyggqcltlvss	401	
YY	142 AKTPKLEEGEFSSEARV-DIVLTGSPAIMSAPGKRYMTCSASSSVSYMMNYOQKSGTS	200	
YY	402 vevsgsgsgsgsgsgvdidqltqspalmasaqgkvtcmtrcasssvsymnygqksgts	461	
YY	201 PKRWIYDTSKLASGVPAHFPGSGSTSTSLTSGMEADDAATYYCOQMSNPFTEGSGTK	260	
YY	462 pkriwtdstksvaagypriyfgsgsgstsyssltsismeaedaalyccqgwsnplltfgatlk	521	
YY	261 LEINRADTAPRGSEQKLISEEDLNHHNNH	291	
YY	522 lel-----khhhhhh	531	
YY	RESULT 7		
YY	AAW82744		
YY	AAW82744 standard; Protein: 281 AA.		
YY	AAW82744;		
YY	10-MAY-1999 (first entry)		
YY	Fusion protein PIC1266/806iscfvtag/his.		
YY	Conjugate; cell targeting; cytotoxic drug; plasmid; fusion protein;		
YY	producing-converting enzyme; cell surface antigen; treatment; cancer;		
YY	inflammation; rheumatoid arthritis; antibody; producing therapy system.		
YY	Synthetic.		
YY	WO9851787-A2.		
YY	19-NOV-1998.		
YY	05-MAY-1998; 98WO-GB01294.		
YY	10-MAY-1997; 97GB-0009421.		
YY	(ZENE) ZENECA LTD.		
YY	Blakey DC, Emery SC;		
YY	WPI: 1999-059700/05.		
YY	N-PSDB: AAV72069.		
YY	New gene construct expressing conjugate of targeting agent and		
YY	producing-converting enzyme - useful for, e.g. targeted production of		
YY	cytotoxic drug in vivo, especially for treatment of cancer		
YY	Example 16; Page 84-85; 100pp: English.		
YY	This sequence is a used in a method for obtaining a novel gene construct		
YY	(A) which expresses, in cells of a mammal, a conjugate (B) of a		
YY	cell-targeting group (I) and a heterologous producing-converting enzyme		
YY	(II), and (B) is directed to leave the cell for selective localisation		
YY	at a cell surface antigen (Ag) recognised by (I). Delivery of (A) to a		
YY	target site, then administration of (III) is used for targeted release of		
YY	cytotoxic drug, specifically for treating cancer but also inflammation		

CC such as rheumatoid arthritis. In situ generation of the targeting
CC antibody increases selectivity, reducing side effects at normal tissue.
CC The method is applicable to any antibody-directed enzyme producing therapy
CC system.

xx
xx Sequence 281 AA;

xx

S0

Query Match 65.8%; Score 1010.5; DB 20; Length 281;
Best Local Similarity 70.4%; Pred. No. 3e-58;
Matches 205; Conservative 20; Mismatches 55; Indels 11; Gaps 3;

QY 1 MKYLPTAAAGLLILIAAPMAOVLOOGSAGELARPASVMSCKASGTYTFRYTMHWK 60
Db 1 mkyllptaaagllililaapamaevqlqsgaelvrgsaavalscdaagnlkdmymhwk 60
QY QRPQGLEMIGWICYNPSRKYTNVNOEKRAKLTLTDKSSSTAYMOLSSLTSEDSAVYYC-A 119
Db qrpqglelmwldwpengdteyapkfrfgrkltdassentaylnhsajlstedtavyychv 120
61 grpegglwmiawidpengdteyapkfrfgrkltdassentaylnhsajlstedtavyychv 120
QY 120 RYDDHYSLDYWGGCTTLTVSSAKTPKLEGEFSEARNVDIVLTOSPAIMSGAPEKVYM 179
Db 121 llygyldmwydgqclsvaavssgggs--g9gsg9g9g9qvltqspaimsaspekvyl 178
QY 180 TCSSASSVSVMNWQOKSGTSPPKRITVTSKLACGVPAHPFGSGSGTSYSLTISGMED 239
Db 179 tcsassvsvlmwivqkqpgtsplklyslstnlaasgvparfs9gsgtsysltistmeed 238
QY 240 AATYACQOMSSNPFFGSGTKLEIRADTAFTGSFOKLISEDLNHHHH 290
Db 239 aatyycqqrstcypltfgaqtktelkr-----eqkliseedlnhhhhh 281

RESULT 8
AAB70842
ID AAB70842 standard; Protein: 330 AA.
XX AAB70842:
AC
XX
DT 25-JUN-2001 (first entry)
XX
DE SNV-env leader/human 7B2-scfv fusion construct.
XX
XX T lymphocyte; antibody; single chain variable antibody; scfv; human;
KW cell-targeting vector; antiviral; cytostatic; immunostimulant; AIDS;
KW gene therapy; immunization; diagnosis; T cell-associated disease; SCID;
KM acquired immune deficiency syndrome; severe combined immune deficiency;
XX T cell lymphoma; fusion construct.
XX
XX Chimeric - Homo sapiens.
OS Chimeric - Spleen necrosis virus.
EH Key Location/Qualifiers
FH Protein 1..45
FT /label= SNV-env_leader
FT 46..330
FT /label= 7B2-scfv
XX
XX DEL946142-Al.
XX
XX 29-MAR-2001.
XX
XX PD 27-SEP-1999; 99DE-1046142.
XX PF 27-SEP-1999; 99DE-1046142.
XX PR
XX (BUND) BUNDESREPUBLIK DEUT PAUL-EHRlich-INST.
PA Cichutek K, Engelstaedtler M;
PI WPI: 2001-246140/26.
XX NR N-PSDB; AAF61511.
XX


```
XX OS Chimeric - Homo sapiens.
XX OS Chimeric - Spleen necrosis virus.
XX FH Key
XX FT Protein
XX FT 1..45
XX FT /label= SNV-env_leader
XX FT 46..309
XX FT /label= K6-scfv
XX PN DE19946142-A1.
XX PD 29-MAR-2001.
XX PF 27-SEP-1999; 99DE-1046142.
XX PR 27-SEP-1999; 99DE-1046142.
XX PA (BUND ) BUNDESREPUBLIK DEUT PAUL-EHRLICH-INST.
XX P Cichutek K, Engelstaedter M;
XX DR WPI; 2001-246140/26.
XX DR N-PSDB; AAF61510.
XX PT Cell-targeting vector selective for T lymphocytes, useful in gene
XX PT therapy of e.g. acquired immune deficiency syndrome, encodes a
XX PT single-chain variable antibody fragment -
XX PS Claim 1; Fig 2; 18pp; German.
XX CC This invention describes a novel cell-targeting vector (A) containing a
XX CC DNA sequence (I) encoding a single-chain variable antibody fragment
XX CC (scfv). The products of the invention have antiviral, cytostatic and
XX CC immunostimulant activity and can be used in gene therapy, immunization
XX CC and diagnosis particularly of T cell-associated diseases, specifically
XX CC acquired immune deficiency syndrome (AIDS), severe combined immune
XX CC deficiency (SCID) or T cell lymphoma. (A) target T cells, independently
XX CC of the CD4 receptor, with high selectivity, 4-5 fold selectivity over
XX CC human B cells, and 1000 fold selectivity over other human cells. A vector
XX CC designated 7A5 encodes a 329 amino acid single-chain variable antibody
XX CC fragment, fully defined in the specification. It was used to transform
XX CC D17 (canine osteosarcoma cells susceptible to spleen necrosis virus
XX CC (SNV)) cells, C8166 (human T lymphocyte) cells, and HeLa (human cervical
XX CC carcinoma) cells. After 48 hours, the cells were stained with x-gal to
XX CC determine transformation. The viral titer (infectious units/ml) was over
XX CC 1 million for D17, 1 million for C8166 but less than 100 for HeLa,
XX CC showing the high selectivity for human T cells. This sequence represents
XX CC the SNV-env leader/human K6-scfv fusion construct used in the
XX CC construction of novel cell targeting vectors described in the invention.
XX SO Sequence 309 AA;

Query Match 62.8%; Score 964.5; DB 22; Length 309;
Best Local Similarity 74.3%; Pred. No. 3.2e-55;
Matches 191; Conservative 16; Mismatches 45; Indels 5; Gaps 2;
```

```
QY 256 GSGTKLEINRADTAPG 272
DB 276 gsgtkleikraaagsg 292

RESULT 11
AAB12563
ID AAB12563 standard; Protein; 239 AA.
XX AC AAB12563;
XX DT 08-NOV-2000 (first entry)
XX DE Anti-TAC sfv protein sequence SEQ ID NO:8.
XX KW Tobacco; extnsin; signal; ext26; ext21; PCR; immunological;
XX KW antibody; mammalian.
XX OS Homo sapiens.
XX PN US6080560-A.
XX PD 27-JUN-2000.
XX PF 25-JUL-1994; 94US-0279772.
XX PR 25-JUL-1994; 94US-0279772.
XX PA (MONS ) MONSANTO CO.
XX PI Russell DR, Fuller JT;
XX DR WPI; 2000-451206/39.
XX DR N-PSDB; AAA60938.
XX PT Producing mammalian antibody in plant cells involves transforming
XX PT tobacco cells with genetic construct comprising a sequence encoding
XX PT single chain antibody and isolating antibody from transformed cells -
XX PS Example; Column 17-18; 11pp; English.
XX CC The present invention describes a method for obtaining a conformationally
XX CC active mammalian antibody from plant cells in a culture involving
XX CC transforming tobacco suspension cells with a DNA construct comprising a
XX CC DNA sequence encoding a secretable mammalian single chain antibody,
XX CC culturing the transgenic cells and isolating the accumulated single
XX CC chain mammalian antibody from the tobacco cells. The method can be used
XX CC for obtaining purified preparations of immunologically active
XX CC conformationally intact mammalian antibody. High yield of commercially
XX CC useful quantities of immunologically active mammalian antibody can be
XX CC recovered by this method. The present sequence represents an anti-TAC sfv
XX CC sequence, which is used in the exemplification of the present invention.
XX SO Sequence 239 AA;

Query Match 62.7%; Score 963.5; DB 21; Length 239;
Best Local Similarity 79.8%; Pred. No. 2.8e-55;
Matches 194; Conservative 7; Mismatches 37; Indels 5; Gaps 2;
```

QY 141 SAKTTPKLEEGEE

CC antibody-PE40 fusion protein is used as an immunotoxin for killing

CC e.g. lymphocytes associated with autoimmune diseases or cancer cells.
 XX Sequence 599 AA;

Query Match 62.3%; Score 957.5; DB 19; Length 599;
 Best Local Similarity 76.7%; Pred. No. 1.8e-54;
 Matches 194; Conservative 8; Mismatches 46; Indels 5; Gaps 2;

QY 23 OVULOOGSGLARPGASVYKMSCKASGTYFTYTHMWKQRPQGLWIGTINPSRGITNY 82
 |||||
 Db 2 qvqlqgsqaelakpgasvkmckasgyfltsymhwkqrp99glwylnpstgyley 61
 |||||
 QY 83 NOKFKDKATLTITDKSSSTAYVQMLSLTSEDSAVYVCARYYDDHSLDYWGQGLTYSSA 142
 |||||
 Db 62 ngkfkdkatltadkssstaymqsltfedaaavycaar---ggvfydw9gqtlitvssg 118
 |||||
 QY 143 KTPPKLEGESEARVDIVLTQSPAIMSASPGEKVTMTCSASSSVSYMNYQOKSGTSPK 202
 |||||
 Db 119 gggs--gggsgggsgqvlvtlqspaimaspgkvtltcasssisymhwfqgkpgtspk 176
 |||||
 QY 203 RWIYDTSKLASGVPAPHRGSGSTYSLTISGMAEDPATYTCOOMSNPTFGSGTKLE 262
 |||||
 Db 177 lwytltsnlaagvparfsgsgtsyltslismaedaatychnqstcylptlftgsgtkle 236
 |||||
 QY 263 INRADTAPTGSEQ 275
 |||||
 Db 237 lkgsllaaltahq 249
 |||||

RESULT 14
 AAY87477
 ID AAY87477 standard; Protein: 599 AA.

XX AAY87477;
 AC 03-JUL-2000 (first entry)
 DT
 XX
 DE Anti-Tac(Fv)-PE40 fusion protein.
 XX
 XX Antibody-toxin fusion protein: single chain antibody; immunotoxin;
 KW Pseudomonas exotoxin; anti-Tac(Fv)-PE40; targeted delivery;
 KM Interleukin-2 receptor; IL-2; helper T-lymphocyte; autoimmune disease;
 KW leukaemia.
 XX
 XX Chimeric - Pseudomonas sp.
 OS Chimeric - Mus sp.
 XX
 XX Key Location/Qualifiers
 FT Region 1..117
 FT Region /note="Anti-Tac heavy chain variable region (VH)"
 FT Region 118..132
 FT Region /note="Linker region"
 FT Region 133..178
 FT Region /note="Anti-Tac light chain variable region (VL)"
 FT Region 179..599
 FT Region /note="PE40"
 FT Misc-difference 392
 FT /note="Encoded by CAC"
 XX
 XX US6051405-A.
 PN 18-APR-2000.
 PD
 XX 08-APR-1992; 92US-0865722.
 PF
 XX 21-APR-1989; 89US-0341361.
 PR 24-SEP-1986; 86US-0911227.
 XX
 XX (USSH) US DEPT HEALTH & HUMAN SERVICES.
 PA (PROT-) PROTEIN DESIGN LABS INC.
 XX
 XX Fitzgerald D, Chaudhary VK, Pastan IH, Queen CL, Waldmann TA;

XX
 DR MPI: 2000-363771/31.
 DR N-PSDB; AAA10398.

PT Construct encoding recombinant scFv-toxin fusion protein to selectively
 PT kill cells bearing antigens or receptors comprises DNA segment
 PS Claim 9; Columns 9-12; 14pp; English.

XX This sequence represents a recombinant single chain antibody
 CC fusion protein anti-Tac(Fv)-PE40, which comprises the heavy and
 CC light chain variable regions (VH and VL) of an anti-Tac antibody and
 CC residues 253-613 of Pseudomonas exotoxin (PE). The anti-Tac single chain
 CC antibody component of the immunotoxin binds to the p55 subunit (Tac
 CC antigen) of the interleukin-2 (IL-2) receptor, which is present in large
 CC amounts on helper T-lymphocytes. This enables the cytotoxic action of
 CC PE40 to be targeted to these and other cell types which express the IL-2
 CC receptor. The recombinant immunotoxin of the invention may be used to
 CC treat a variety of autoimmune diseases, including graft-versus-host
 CC disease, organ transplant rejection, type I diabetes, multiple
 CC sclerosis, rheumatoid arthritis, systemic lupus erythematosus and
 CC myasthenia gravis. It may also be used in the treatment of leukaemia,
 CC and may be used in vitro for the elimination of harmful cells from bone
 CC marrow before transplant. Anti-Tac(Fv)-PE40 is more active on a molar
 CC basis than anti-Tac antibody chemically conjugated to full-length PE
 CC (anti-Tac-PE) or to PE40 (anti-Tac-PE40).

SO Sequence 599 AA;

Query Match 62.3%; Score 957.5; DB 21; Length 599;
 Best Local Similarity 76.7%; Pred. No. 1.8e-54;
 Matches 194; Conservative 8; Mismatches 46; Indels 5; Gaps 2;

QY 23 OVULOOGSGLARPGASVYKMSCKASGTYFTYTHMWKQRPQGLWIGTINPSRGITNY 82
 |||||
 Db 2 qvqlqgsqaelakpgasvkmckasgyfltsymhwkqrp99glwylnpstgyley 61
 |||||
 QY 83 NOKFKDKATLTITDKSSSTAYVQMLSLTSEDSAVYVCARYYDDHSLDYWGQGLTYSSA 142
 |||||
 Db 62 ngkfkdkatltadkssstaymqsltfedaaavycaar---ggvfydw9gqtlitvssg 118
 |||||
 QY 143 KTPPKLEGESEARVDIVLTQSPAIMSASPGEKVTMTCSASSSVSYMNYQOKSGTSPK 202
 |||||
 Db 119 gggs--gggsgggsgqvlvtlqspaimaspgkvtltcasssisymhwfqgkpgtspk 176
 |||||
 QY 203 RWIYDTSKLASGVPAPHRGSGSTYSLTISGMAEDPATYTCOOMSNPTFGSGTKLE 262
 |||||
 Db 177 lwytltsnlaagvparfsgsgtsyltslismaedaatychnqstcylptlftgsgtkle 236
 |||||
 QY 263 INRADTAPTGSEQ 275
 |||||
 Db 237 lkgsllaaltahq 249
 |||||

RESULT 15
 AAM95462
 ID AAM95462 standard; Protein: 599 AA.

XX AAM95462;
 AC 23-MAR-1999 (first entry)
 DT
 XX
 DE Anti-Tac(Fv)-PE40 fusion protein.
 XX
 XX Antibody toxin fusion protein: receptor; immunotoxin; scFv; PE40;
 KW fusion protein; single-chain Fv; Pseudomonas; exotoxin; translocating;
 KM ADP-ribosylating; autoimmune disease; cancer; anti-Tac(Fv)-PE40.
 XX
 XX Synthetic.
 OS Pseudomonas sp.
 OS
 XX
 XX Key Location/Qualifiers

```

FT Misc-difference 27 /note= "encoded by GCG"
FT Misc-difference 92 /note= "encoded by TGT"
FT Misc-difference 99 /note= "encoded by ACA"
FT Misc-difference 105 /note= "encoded by CAC"
FT Misc-difference 118 /note= "encoded by GCA"
FT Misc-difference 124 /note= "encoded by GCT"
FT Misc-difference 148 /note= "encoded by GCG"
FT Misc-difference 175 /note= "encoded by GCG"
FT Misc-difference 177 /note= "encoded by GCG"
FT Misc-difference 197 /note= "encoded by GTC"
FT Misc-difference 202 /note= "encoded by GCA"
FT Misc-difference 282 /note= "encoded by TAG"
FT Misc-difference 298 /note= "encoded by GGC"
FT Misc-difference 307 /note= "encoded by ATG"
FT Misc-difference 308 /note= "encoded by GGC"
FT Misc-difference 327 /note= "encoded by AGC"
FT Misc-difference 348 /note= "encoded by AGG"
FT Misc-difference 377 /note= "encoded by CCG"
FT Misc-difference 394 /note= "encoded by GAC"
FT Misc-difference 401 /note= "encoded by ACC"
FT Misc-difference 436 /note= "encoded by CAC"
FT Misc-difference 440 /note= "encoded by ATG"
FT Misc-difference 497 /note= "encoded by TAT"
FT Misc-difference 509 /note= "encoded by TTG"
FT Misc-difference 545 /note= "encoded by CGG"
FT Misc-difference 548 /note= "encoded by GCG"
FT Misc-difference 562 /note= "encoded by CAG"
FT Misc-difference /note= "encoded by GGC"

US5863745-A.
26-JAN-1999.
XX
XX
XX 05-JUN-1995; 95US-0461825.
XX
XX 21-APR-1989; 89US-0341361.
XX 24-SEP-1986; 86US-0911227.
XX 08-APR-1992; 92US-0865722.
XX 05-JUN-1995; 95US-0461825.
XX
XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX Chaudhary VK, Fitzgerald DJ, Pastan IH, Queen CL,
XX Waldmann TA;
XX WPI; 1999-131300/11.
XX N-PSDB; AAX00812.
DR

```

```

XX
XX Killing cells with immunotoxin - comprising recombinant fusion
XX protein of antibody Fv fragment and truncated Pseudomonas exotoxin
XX
XX Disclosures: Columns 9-12; 14pp; English.
XX
XX The invention relates to recombinant antibody toxin fusion proteins which
XX selectively kill cells bearing appropriate antigens or receptors. Killing
XX cells with immunotoxin comprises contacting cells targeted to be killed
XX with an antibody-PE40 recombinant fusion protein, where the antibody is a
XX single-chain Fv fragment (scFv) and the PE40 is a Pseudomonas exotoxin
XX (PE) fragment that lacks amino acids 1-252 and which has at least the
XX translocating and ADP-ribosylating activity of PE, where the cells
XX targeted to be killed have receptors or antigens to which the antibody
XX binds and the fusion protein has lower toxicity to cells that lack such
XX receptors or antigens. The method is useful for treating autoimmune
XX diseases or cancer. The present sequence represents an anti-Tac(Fv)-PE40
XX fusion protein encoded by a Vm-VL-Linker-PE40 DNA segment.
XX
XX Sequence 599 AA:
SQ

```

```

Query Match 62.0%, Score 952.5; DB 20; Length 599;
Best Local Similarity 76.3%; Pred. No. 3.9e-54;
Matches 193; Conservative 8; Mismatches 47; Indels 5; Gaps 2;

QY 23 QYQLOQSGAEIARPGASVKNMCKASGYTFRTYTMWVKORQGLEWIGYINPSKGTNY 82
   |||||
DB 2 9vqlqgsgealokpgsvkmsckasgytlfsymwvqrpqglewlygimpstgyley 61
   |||||

QY 83 NQFKDKATLTFTDKSSSTAYWQLSLTSEDSAVYVCARYDDHYSLDYWGQGTTLTVSSA 142
   |||||
DB 62 ngfkdkatltadkssstaaymlsltfedaavyycaar---ggvfyfwggqtlvtvssg 118
   |||||

QY 143 KTPPKLEGCESEARVDIVLQSPAIMSASPEKVTMTCSASSSVSYNMWYQOKSGTSPK 202
   |||||
DB 119 gggs--gggsgsggsgqivltqspaimaspekvltlcsassisyhwhfqqkpslpk 176
   |||||

QY 203 RMIYPTSKLASGVPAHFPGSGSGTSTISGMEDADATYYCOQMSNPTFFSGTGLE 262
   |||||
DB 177 lwiyltslnlasvparpatsgsgtsysltlsmeeadaalychqrslypltfsgtkle 236
   |||||

QY 263 INRADTAPTSSEQ 275
   |||||
DB 237 lkgsllaaltahq 249
   |||||

```

Search completed: March 5, 2002, 17:05:34
Job time: 32 sec

Wed Mar 6 05:48:53 2002

us-09-424-705-2.rag

Page 11

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 5, 2002, 17:05:02 ; Search time 16.04 Seconds
(without alignments)
1381.970 Million cell updates/sec

Title: US-09-424-705-2

Perfect score: 1536
Sequence: 1 MKYLLPTAAGLLLAQPA.....GSEOKLSEEDINSHHHHH 291

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

1: PIR_68:*
2: PIR1:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	857.5	55.8	268	A56446	Ig heavy chain V r
2	671	43.7	249	S41374	single chain Fv an
3	612	39.8	233	JC5322	p53 specific, singl
4	545	35.5	144	B30502	Ig heavy chain V r
5	540.5	35.2	139	B50024	Ig heavy chain pre
6	531	34.6	235	S25058	Ig kappa chain - m
7	527	34.3	287	PC4402	pelb leader/Ig hea
8	521.5	34.0	246	S38950	Ig gamma chain - m
9	521.5	34.0	446	S40295	Ig gamma-2a chain
10	519	33.8	107	A30562	Ig kappa chain V r
11	519	33.8	474	G2MS11	Ig gamma-2b chain
12	514	33.5	107	B30562	Ig kappa chain V r
13	514	33.5	130	J10079	Ig kappa chain pre
14	513.5	33.4	120	S41394	Ig heavy chain V r
15	512.5	33.4	140	PH1482	Ig heavy chain V r
16	511.5	33.3	139	A27609	Ig heavy chain pre
17	508.5	33.1	140	HVMSG7	Ig heavy chain V r
18	508	33.1	103	S29591	Ig kappa chain V r
19	506.5	33.0	120	B22769	Ig heavy chain V r
20	504	32.8	104	B49049	Ig kappa chain V r
21	503.5	32.8	122	S24287	Ig heavy chain V r
22	503.5	32.8	131	S66537	Ig heavy chain V r
23	503.5	32.8	139	MMMS18	Ig heavy chain pre
24	501.5	32.6	469	S37483	Ig gamma-2a chain
25	501	32.6	106	PS0071	Ig kappa chain V r
26	499.5	32.5	120	S25175	Ig heavy chain V r
27	499.5	32.5	140	PH1489	Ig heavy chain V r
28	497.5	32.4	108	G30560	Ig kappa chain V r
29	497	32.4	118	S38565	Ig heavy chain V r

30	496	32.3	116	2	S55442	Ig heavy chain V r
31	496	32.3	121	2	A26405	Ig heavy chain V r
32	494.5	32.2	116	2	S53751	antibody Fab Jcl 1
33	494	32.2	107	2	S11118	Ig kappa chain V r
34	494	32.2	122	2	S20643	Ig heavy chain V r
35	493.5	32.1	140	2	PH1484	Ig heavy chain V r
36	493	32.1	107	2	PR0406	Ig kappa chain V r
37	492.5	32.1	151	2	PL0011	Ig heavy chain pre
38	492	32.0	123	2	E48677	Ig heavy chain V-D
39	491	32.0	107	2	PC4405	Ig kappa chain V r
40	491	32.0	107	2	S11119	Ig kappa chain V r
41	491	32.0	138	2	E32513	Ig heavy chain pre
42	490.5	31.9	123	2	S60067	Ig heavy chain V r
43	490.5	31.9	140	2	PH1488	Ig heavy chain V r
44	490	31.9	123	2	S20646	Ig heavy chain V r
45	489.5	31.9	140	2	PH1498	Ig heavy chain V r

ALIGNMENTS

RESULT 1
A56446
Ig heavy chain V region (3H-3H scFv) - mouse (strain BALB/C)
C:Species: Mus musculus (house mouse)
C:Date: 19-Jan-1996 #sequence_revision 19-Jan-1996 #text_change 16-Aug-1996
C:Accession: A56446
R:Tang, P.M.; Foltz, L.A.; Mahoney, W.C.; Schueler, P.A.
J. Biol. Chem. 270, 7829-7835, 1995
A:Title: A high affinity digoxin-binding protein displayed on M13 is functionally ide
A:Reference number: A56446; MUID:95229583
A:Accession: A56446
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-268 <TAN>
A:Cross-references: GB:020617
C:Keywords: heterotetramer; immunoglobulin

Query Match 55.8%; Score 857.5; DB 2; Length 268;
Best Local Similarity 65.3%; Pred. No. 1.1e-51;
Matches 175; Conservative 25; Mismatches 61; Indels 7; Gaps 3;
Query 21 MAOYLOQSGAELAPGASVMSCKASGYTFRTYTHMWKORPGGLEWIGYINPSRGYT 80
Db 1 MAOYVLOSGAELAPGASVMSCKASGYTFRTYTHMWKORPGGLEWIGRIAPANGIT 60
Query 81 NYNOKFKDKATLTDDSSSTAYMQLSLTSEDSAVYCARYYDDHYSLDWGQGTTFVS 140
Db 61 KYDPFKQKATLTADTSSNTAYLQLSLTSEDTAVYCYASYLTRYE-NYMGQGTTFVS 119
Query 141 SAKTTPKLEDEGFSEARADIVLTQSPALMSASPGKRYMTCSASSSVYMMYQOKSGTS 200
Db 120 SGGGGS--GGDSSGGGSDIELTQSPALMSASLGKRYMTCSRASSVNFYIYQOKSDAS 177
Query 201 PKRWIYDTSKLASCVPAPFRSGSGSTSYSLTISGAEADATYYCOQSSNPFPGSGTK 260
Db 178 PKLWYITSHLPKVPAPFRSGSGSGNSYSLTISSHGEGDAATYYCOQRTSSPFTGSGTK 237
Query 261 LEINRADT---APTGSOKLISEEDLN 284
Db 238 LEIKRSAAHHHHHNGAABOKLISEEDLN 265
RESULT 2
S41374
single chain Fv antibody - mouse
C:Species: Mus musculus (house mouse)
C:Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 06-Jan-1995
C:Accession: S41374
R:Arasanko, O.; Weller, E.W.; Muentz, K.; Conrad, U.
submitted to the EMBL Data Library, January 1994
A:Description: Construction and functional characterization of a single chain Fv anti

A:Reference number: S41374
 A:Accession: S41374
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-249 <ART>
 A:Cross-references: EMBL:Z29480

Query Match 43.7%; Score 671; DB 2; Length 249;
 Best Local Similarity 56.0%; Pred. No. 5.3e-39;
 Matches 140; Conservative 29; Mismatches 73; Indels 8; Gaps 2;

23 OVOLOOQSGAELARPASVYKMSKASGYTFTRYTMHWKORPGGLEWIGYINPSRGYTN 82
 1 OVOLOOQSGAELVPGASVYKMSKASGYTFTRYTMHWKORPEKGLWIRIAPASGNVX 60
 83 NOKFKDKATLTDDKSSSTAYMQLSLTSEDSAVYICARYDDHYSLDYWGQGTTLTVSSA 142
 61 VRFQDKATLTADTSSNTAYLQLSLTSEDTAVYICARDTLTYSLGWGQGSTVTVSSR 120
 143 KTTPLKEGEFSEARVDIYLTQSPAIMSASPEKVTMTCSASSV-----SYMNNYOOK 196
 121 GGG--GGGSGGGGSDILTQSPSPVYIPESVVISCRSSKSLYSGDSTLFWFLOR 178
 197 SGTSPKRWIYDTSKLASGVPAHFRSGSGTSLTISGMEADPAATYVCOQMSNPFPG 256
 179 PQSGPQLLIRMSNLASGVDPDRFSGSGSTFTLRISRYEADVDYVYCMQHREYDLTFG 238
 257 SGTKLEINRA 266
 239 AGTKLELKRA 248

RESULT 3

p53 specific single-chain antibody pab421 - human
 C:Species: Homo sapiens (man)
 C:Date: 15-May-1997 #sequence_revision 15-May-1997 #text_change 18-Jul-1997
 C:Accession: J05322

R:Jannot, C.B.; Hynes, N.E.
 Biochem. Biophys. Res. Commun. 230, 242-246, 1997
 A:Title: Characterization of scFv-421, a single-chain antibody targeted to p53.
 A:Reference number: J05322; MUID:97168950
 A:Accession: J05322
 A:Molecule type: mRNA
 A:Residues: 1-233 <JAN>
 A:Experimental source: hybridoma cell
 A:Comment: This protein specifically binds the tumor suppressor protein p53. It restores

Query Match 39.8%; Score 612; DB 2; Length 233;
 Best Local Similarity 53.6%; Pred. No. 5.3e-35;
 Matches 133; Conservative 26; Mismatches 63; Indels 26; Gaps 5;

27 OOSGAELARPASVYKMSKASGYTFTRYTMHWKORPGGLEWIGYINPSRGYTNOKF 86
 1 OESGAELVRSGASVYKMSKASGYTFTRYTMHWKORPEKGLWIRIAPASGNVX 60
 87 KDKATLTDDKSSSTAYMQLSLTSEDSAVYICARYDDHYSLDYWGQGTTLTVSSAKTTP 146
 61 GVKATMTADTSSNTAYLQLSLTSEDTAVYIC-----NAGMDYWGQGTTVTVSSG---- 110
 147 KLEEGEFSEARV-----DIVLTQSPAIMSASPEKVTMTCSASSV-----SYMNNYOOK 195
 111 ---GGSGSGRASGGGSDILTQSPASVLAIGORATISCRASSVSTISGYSYHMMNQ 166
 196 KSGTSPKRWIYDTSKLASGVPAHFRSGSGTSLTISGMEADPAATYVCOQMSNPFPG 255
 167 KGPQPRLLIYVSNLESGVPARFSGSGGDTFTLIRHVEEDATYVYCOHIRELTRE 226
 256 GSGTKLEI 263
 227 G-GTKLEI 233

RESULT 4

Ig heavy chain V region (D444) - mouse
 B30502
 C:Species: Mus musculus (house mouse)
 C:Date: 03-Nov-1988 #sequence_revision 03-Nov-1988 #text_change 21-Jan-2000
 C:Accession: B30502
 R:Eilat, D.; Webster, D.M.; Rees, A.R.
 J. Immunol. 141, 1745-1753, 1988
 A:Title: V region sequences of anti-DNA and anti-RNA autoantibodies from NZB/NZW F-1
 A:Reference number: A30502; MUID:88315787
 A:Accession: B30502
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-144 <EIL>
 C:Superfamily: immunoglobulin V region; immunoglobulin homology
 C:Keywords: heterotetramer; immunoglobulin
 F:15-98/Domain: immunoglobulin homology <IMM>

Query Match 35.5%; Score 545; DB 2; Length 144;
 Best Local Similarity 65.5%; Pred. No. 1.1e-30;
 Matches 108; Conservative 11; Mismatches 22; Indels 24; Gaps 2;

23 OVOLOOQSGAELARPASVYKMSKASGYTFTRYTMHWKORPGGLEWIGYINPSRGYTN 82
 1 EVOLOOQSGPELVKPGASVYKMSKASGYTFTRYTMHWKORPGGLEWIGYINPYDGTXY 60
 83 NOKFKDKATLTDDKSSSTAYMQLSLTSEDSAVYICARYDDHYSLDYWGQGTTLTVSSA 142
 61 NEKFKGKATLTDDKSSSTAYMQLSLTSEDSAVYICAR-----CGFDYWGQGTTLT---- 111
 143 KTTPLKEGEFSEARVDIYLTQSPAIMSASPEKVTMTCSASSV 187
 112 -----VDILTQSPAILSVSPERVSFSCRASQSI 141

RESULT 5

Ig heavy chain precursor V region (6A4) - mouse
 PS0024
 C:Species: Mus musculus (house mouse)
 C:Date: 07-Jun-1990 #sequence_revision 07-Jun-1990 #text_change 22-Nov-1996
 C:Accession: PS0024
 R:Marget, M.; Eckhardt, A.; Ehret, W.; von Specht, B.U.; Duchene, M.; Domdey, H.
 Gene 74, 335-345, 1988
 A:Title: Cloning and characterization of cDNAs coding for the heavy and light chains
 A:Reference number: PS0023; MUID:89232725
 A:Accession: PS0024
 A:Molecule type: mRNA
 A:Residues: 1-139 <MAR>
 A:Experimental source: strain BALB/c
 C:Comment: This chain is obtained from an IgG2a monoclonal antibody against Pseudomonas

C:Superfamily: immunoglobulin V region; immunoglobulin homology
 C:Keywords: heterotetramer; immunoglobulin; pyroglytamic acid
 F:1-19/Domain: signal sequence #status predicted <SIG>
 F:20-113/Domain: Ig heavy chain V region #status predicted <IGV>
 F:34-117/Domain: immunoglobulin homology <IMM>
 F:20/Modified site: pyroglutamic acid (Gln) (in mature form) #status predicted

Query Match 35.2%; Score 540.5; DB 2; Length 139;
 Best Local Similarity 84.3%; Pred. No. 2.2e-30;
 Matches 102; Conservative 8; Mismatches 10; Indels 1; Gaps 1;

22 AOVLOOQSGAELARPASVYKMSKASGYTFTRYTMHWKORPGGLEWIGYINPSRGYTN 81
 19 SOVLOOQSGAELARPASVYKMSKASGYTFTRYTMHWKORPGGLEWIGYINPTGYTE 78
 82 VNOKFKDKATLTDDKSSSTAYMQLSLTSEDSAVYICAR-YDDHYSLDYWGQGTTLTVS 140
 79 YNOKFKDKATLTADKSSSTAYMQLSLTSEDSAVYICRTSYNYEGAMDYWGQGTSTVS 138

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 5, 2002, 17:05:57 ; Search time 12.93 Seconds

(without alignments)
825,171 Million cell updates/sec

Title: US-09-424-705-2

Perfect score: 1536

Sequence: 1 MKYLLPTAAAGLLLLAAQPA.....GSEQKLISEEDLNHHHHH 291

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 100059 segs, 36664827 residues

Post-processing: Minimum Match 0%

Database : SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	508.5	33.1	140	1	HV02_MOUSE
2	508	33.1	120	1	HV03_MOUSE
3	503.5	32.8	139	1	HV07_MOUSE
4	500	32.6	107	1	KV6F_MOUSE
5	497	32.4	107	1	KV6I_MOUSE
6	495	32.2	107	1	KV6I_MOUSE
7	484	32.2	107	1	KV6H_MOUSE
8	481	32.0	107	1	KV6G_MOUSE
9	482.5	31.4	137	1	HV11_MOUSE
10	463	30.1	138	1	HV48_MOUSE
11	462.5	30.1	120	1	HV50_MOUSE
12	462	30.1	117	1	HV12_MOUSE
13	457	29.8	117	1	HV13_MOUSE
14	446	29.0	107	1	KV6E_MOUSE
15	446	29.0	121	1	HV01_MOUSE
16	443	28.8	107	1	KV6A_MOUSE
17	443	28.8	117	1	HV04_MOUSE
18	442	28.8	107	1	KV6D_MOUSE
19	441	28.7	107	1	KV6C_MOUSE
20	440	28.6	136	1	HV15_MOUSE
21	436.5	28.4	118	1	HV51_MOUSE
22	436	28.4	117	1	HV05_MOUSE
23	434	28.3	117	1	HV06_MOUSE
24	427	27.8	108	1	KV6K_MOUSE
25	426	27.7	117	1	HV09_MOUSE
26	424	27.6	107	1	KV6E_MOUSE
27	422	27.5	117	1	HV49_MOUSE
28	412.5	26.9	117	1	HV10_MOUSE
29	411.5	26.8	117	1	HV52_HUMAN
30	410.5	26.7	117	1	HV14_MOUSE
31	407	26.5	129	1	KV4A_MOUSE
32	407	26.5	129	1	KV4A_MOUSE
33	389	25.3	117	1	HV1B_HUMAN

34	380	24.7	117	1	HV1G_HUMAN	P23083	homo sapien
35	373.5	24.4	114	1	HV02_MOUSE	P01741	mus musc
36	349.5	22.8	111	1	KV3M_MOUSE	P01655	mus musc
37	346.5	22.6	108	1	KV1V_HUMAN	P04430	homo sapien
38	345.5	22.5	111	1	KV3O_MOUSE	P01657	mus musc
39	343.5	22.4	108	1	KV1O_HUMAN	P01607	homo sapien
40	342.5	22.3	111	1	KV3L_MOUSE	P01654	mus musc
41	339.5	22.1	111	1	KV3Q_MOUSE	P01594	homo sapien
42	338.5	22.0	108	1	KV1B_HUMAN	P01666	mus musc
43	338.5	22.0	111	1	KV3N_MOUSE	P01608	homo sapien
44	337.5	22.0	108	1	KV1P_HUMAN	P01807	mus musc
45	335.5	21.8	119	1	HV37_MOUSE		

ALIGNMENTS

RESULT	ID	HY02_MOUSE	STANDARD	PRT	140 AA
AC	P01746	21-JUL-1986 (Rel. 01, Created)			
DT	21-JUL-1986 (Rel. 01, Last sequence update)				
DT	15-JUL-1999 (Rel. 38, Last annotation update)				
DE	IG HEAVY CHAIN V REGION 9367 PRECURSOR.				
OS	Mus musculus (Mouse).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murine; Mus.				
OX	NCBI_TaxID=10090;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN-A/J;				
RX	MEDLINE=82152818; PubMed=6801765;				
RA	Sims J., Rabbits T.H., Estess P., Slaughter C., Tucker P.W.,				
RA	Capra J.D.,				
RT	"Somatic mutation in genes for the variable portion of the				
RT	immunoglobulin heavy chain."				
RL	Science 216:309-311(1982).				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/				
CC	or send an email to license@sib-sib.ch).				
CC	-----				
CC	EMBL: J00493; AAA8128.1; -				
DR	PIR: A02028; HVM5G7.				
DR	InterPro: IPR003006; IG_MHC.				
DR	InterPro: IPR003596; IG_V.				
DR	Pfam: PF00047; Ig_V.				
DR	SMART: SM00406; IgV_1.				
KW	Immunoglobulin V region; Antitarsionate antibody; Hybridoma; Signal.				
FT	CHAIN 1 19				
FT	CHAIN 20 140				
FT	NON TER 140 140				
FT	SEQUENCE 140 AA; 15514 MW; 25AACBBE31DA5CE8 CRC64;				
QY	12 LLLLAAPAM-AQVOLOOSGAEELARPGASVYKMSCKASGYTFTRTYTMVAVQRPQGLEMI 70				
QY	8 LFLSLVTAGVHSEVOLOOOSGAEVLRAGSSVYKMSCKASGYTFTSYGIMVAVQRPQGLEMI 67				
QY	71 GYINPSRGYTYNOKFKDKATILTTDKSSSTAYKMLSSSTEDSAVYYCAR--YYDDHYSL 128				
QY	68 GYINPGMGTYINYNKFRGKRTITLVDRKSSSTAYKMLRSLTSDSAVYCARSHYTGGSYDF 127				
QY	129 DYWGQGTTLTVSS 141				

Query Match 33.1%; Score 508.5; DB 1; Length 140;
Best local similarity 74.4%; Pred. No. 2e-33;
Matches 99; Conservative 9; Mismatches 22; Indels 3; Gaps 2;

```

CC CC MAKING ANTIBODIES TO THE HAPTEN (4-HYDROXY-3-NITROPHENYL)ACETYL
CC CC (NIB ANTIBODIES).
CC CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC CC at the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC -----
CC CC EMBL: J00529; AAA38170.1; -.
CC CC PIR: A02034; MHMS18.
CC CC InterPro: IPR003006; IG_MHC.
CC CC InterPro: IPR003596; IG_V.
CC CC Pfam: PF00047; Ig, 1.
CC CC SMART: SM00406; Igv, 1.
CC CC Immunoglobulin V region; Signal.
CC CC FT SIGNAL 1 19
CC CC FT CHAIN 20 139 IG HEAVY CHAIN V REGION B1-8/186-2.
CC CC FT DOMAIN 20 49 FRAMEWORK 1.
CC CC FT DOMAIN 50 54 COMPLEMENTARITY-DETERMINING 1.
CC CC FT DOMAIN 55 58 FRAMEWORK 2.
CC CC FT DOMAIN 69 85 COMPLEMENTARITY-DETERMINING 2.
CC CC FT DOMAIN 86 117 FRAMEWORK 3.
CC CC FT DOMAIN 118 124 D SEGMENT.
CC CC FT DOMAIN 125 139 JH2 SEGMENT.
CC CC FT DISULFD 41 115 BY SIMILARITY.
CC CC FT NON_TER 139 139
CC CC SQ SEQUENCE 139 AA; 15419 MW; 1B57DD4FD0C9F465 CRC64;
CC CC -----
CC CC Query Match 32.8%; Score 503.5; DB 1; Length 139;
CC CC Best Local Similarity 74.8%; Pred. No. 4.8e-33;
CC CC Matches 101; Conservative 9; Mismatches 18; Indels 7; Gaps 3;
CC CC -----
OY 12 LILLAAQF-AMAAQVQLDSGAEELARPGASVSKASGYFTFRTYTHMVAKRPDQGLEW 69
DB 7 MLEFLAATATGVSHSQVQLDQPGAEELVKGASVSKLSCKASGATFTSTYMWVAKRPDQGLEW 66
OY 70 IGYINPSKGYTYNYNOKFEDKATLFTDKSSSTAYVMQLSSLTSEDSAVVYICARYDDHYS-- 127
DB 67 IGRIDPNNGGRTKYNKREKSKATLVYDKRPSSTAYVMQLSSLTSEDSAVVYICARY-- 124
OY 128 -LDYWGQGTFLTVSS 141
DB 125 YEDYWGQGTFLTVSS 139
CC CC -----
CC CC RESULT 4
CC CC KVF_MOUSE STANDARD; PRT; 107 AA.
CC CC ID KVF_MOUSE PO4940:
CC CC DT 13-AUG-1987 (Rel. 05, Created)
CC CC DT 13-AUG-1987 (Rel. 05, Last sequence update)
CC CC DT 15-JUL-1999 (Rel. 38, Last annotation update)
CC CC DE IG KAPPA CHAIN V-VI REGION NQ2-17.4.1.
CC CC OS Mus musculus (Mouse).
CC CC CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC CC OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
CC CC OX NCBI_TaxID=10090;
CC CC RN
CC CC RP SEQUENCE FROM N.A.
CC CC RX MEDLINE=83271467; PubMed=6877353;
CC CC RA Kaartinen M., Griffiths G.M., Markham A.F., Milstein C.;
CC CC RT "mRNA sequences define an unusually restricted igc response to 2-
CC CC RL phenylloxazalone and its early diversification.";
CC CC Nature 304:320-324(1983).
CC CC -I- FUNCTION: ANTI-2-PHENYL OXAZOLONE (PHOX) ANTIBODY.
CC CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its

```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 5, 2002, 17:05:37 ; Search time 24.4 Seconds

(without alignments)
1744.477 Million cell updates/sec

Title: US-09-424-705-2

Perfect score: 1336
Sequence: 1 MKYLLPTAAAGLLLAQPA.....GSEOKLISEEDLNHHNNH 291

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPREMBL_17:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	690	44.9	298	11 Q9QYF0	Q9QYF0 mus musculu
2	531	34.6	119	5 Q9GYZ2	Q9GYZ2 schistosoma
3	510	33.2	473	11 Q991J25	Q991J25 mus musculu
4	503	32.7	473	11 Q9DBL4	Q9DBL4 mus musculu
5	498	32.4	463	11 Q991C4	Q991C4 mus musculu
6	480.5	31.3	468	11 Q991J31	Q991J31 mus musculu
7	477	31.1	117	11 Q9QXF0	Q9QXF0 mus musculu
8	477	30.4	109	11 Q9QXK9	Q9QXK9 mus musculu
9	467.5	30.1	109	11 Q9JL75	Q9JL75 mus musculu
10	462.5	29.6	114	11 Q921C4	Q921C4 mus musculu
11	454.5	28.9	106	5 Q9U410	Q9U410 schistosoma
12	444	28.7	110	11 Q9JL77	Q9JL77 mus musculu
13	440.5	27.0	117	11 Q921C6	Q921C6 mus musculu
14	415	27.0	500	4 Q9BRV0	Q9BRV0 homo sapien
15	413.5	26.9	124	4 Q9UL92	Q9UL92 homo sapien
16	412	26.8	484	11 Q991A6	Q991A6 mus musculu
17	408	26.6	119	4 Q9UL94	Q9UL94 homo sapien
18	404.5	26.3	97	11 Q9JL76	Q9JL76 mus musculu

20	402	26.2	111	11 Q9DB8	Q9DB8 mus musculu
21	398	25.9	125	4 Q9UL95	Q9UL95 homo sapien
22	384.5	25.0	110	11 Q9JL83	Q9JL83 mus musculu
23	377	24.5	150	4 Q9Y298	Q9Y298 homo sapien
24	375	24.4	109	11 Q9JL85	Q9JL85 mus musculu
25	369.5	24.1	116	4 Q9UL89	Q9UL89 homo sapien
26	363.5	23.7	214	11 Q9JL85	Q9JL85 mus musculu
27	358.5	23.3	101	11 Q9JL78	Q9JL78 mus musculu
28	356.5	23.2	157	4 Q95978	Q95978 homo sapien
29	340	22.1	109	4 Q9UL78	Q9UL78 homo sapien
30	340	22.1	487	11 Q99KA4	Q99KA4 mus musculu
31	335.5	21.8	108	4 Q9UL77	Q9UL77 homo sapien
32	335.5	21.8	108	4 Q9UL70	Q9UL70 homo sapien
33	335	21.8	107	4 Q9UL81	Q9UL81 homo sapien
34	330	21.5	597	4 Q9BOB8	Q9BOB8 homo sapien
35	329.5	21.5	108	4 Q9UL79	Q9UL79 homo sapien
36	329	21.4	597	4 Q9BU10	Q9BU10 homo sapien
37	323.5	21.1	108	4 Q9UL83	Q9UL83 homo sapien
38	322	21.0	147	4 Q9Y509	Q9Y509 homo sapien
39	321	20.9	109	4 Q9UL86	Q9UL86 homo sapien
40	319	20.8	113	4 Q9UL90	Q9UL90 homo sapien
41	318	20.7	437	11 Q9JL44	Q9JL44 mus musculu
42	317	20.6	109	4 Q9UL85	Q9UL85 homo sapien
43	316	20.6	238	11 Q99M37	Q99M37 mus musculu
44	315.5	20.5	112	4 Q9HCC1	Q9HCC1 homo sapien
45	313	20.4	120	4 Q9BUA1	Q9BUA1 homo sapien

ALIGNMENTS

RESULT	1	PRELIMINARY:	PRT:	298 AA.
ID	Q9QYF0			
AC	Q9QYF0			
DT	01-MAY-2000 (TREMBLrel. 13, Created)			
DT	01-MAY-2000 (TREMBLrel. 13, Last sequence update)			
DT	01-JUN-2001 (TREMBLrel. 17, Last annotation update)			
DE	CN 8 SCFV.			
GN	CN 8			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.			
OX	NCBI_Taxid:10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-BALB/C; TISSUE-SPLEEN;			
RA	Shinozaki N.; Demura T.; Fukuda H.;			
RT	"Isolation of a novel type of vascular cell wall-specific monoclonal antibody recognizing a cell polarity using a phage display subtraction method."			
RT	Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.			
CC	-1- SIMILARITY: TO IMMUNOGLOBULIN AND MAJOR HISTOCOMPATIBILITY COMPLEX DOMAIN.			
CC	EMBL; AB036341; BAA8633.1; -			
DR	HSSP; P01607; IREI.			
DR	InterPro; IPR003006; IG_MHC.			
DR	InterPro; IPR003596; IG_V.			
DR	Pfam; PF00047; IG_2.			
DR	SMART; SM00406; IGV_2.			
SO	SEQUENCE 298 AA; 31867 MW; EOF96B8A17004317 CRC64;			

Query Match 44.9%; Score 690; DB 11; Length 298;
Best Local Similarity 54.7%; Pred. No. 4.9e-50;
Matches 139; Conservative 35; Mismatches 72; Indels 8; Gaps 5;
QY 16 AAGPAAQVQLDQSGAELARPGASVKNCSKASGYTFRTYTHMWKQPGGLEWIGYINP 75

```

Db 33 AAOVAMQVKKLOQSGGGLVKKPGGSLKLSGASGSDPSRYSWMSWVQAPGKGLWIGETINP 92
Qy 76 SRGTNNQKRRKATLTITDSSSTAYWQSLTSEDSAVVYCAR--YYDDHYSIDYMGQ 133
Db 93 DSSTINTYPSLKDKFTISRDNAKNFLYLQMSKVRESDTALYCAHASY--YWGQ 149
Qy 134 GTTLVSSAKTTPKLESEFSEARVDIVLTQSPAIMSAPGKRYMTCSASSV--SYNMW 192
Db 150 GTTYVSSGGGGS--GGGGSGGGGSDIELTQSPALSSAVGFTVITCRASGNINYLAW 207
Qy 193 YQKSGSTSPKRWYDTSKLAGVPAHFRGSGSGTSYSLTISGMEADATYYCQGMSSNP 252
Db 208 YQKQKSGPQLLYNAKTLADGVPSRFGSGSGTGYSLKINSLQEDFGSYCQHFTTTP 267
Qy 253 FTSGSGTKLEINRA 266
Db 268 YTFGGGTKLEIKRA 281

JUL 2
09GYZ2 PRELIMINARY; PRT; 119 AA.
AC 09GYZ2;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE MONOCLOML ANTI-IDIOTYPIC ANTIBODY NP30 HEAVY CHAIN VARIABLE REGION
DE (FRAGMENT).
OS Schistosoma japonicum (Blood fluke).
OC Eukaryota; Metazoa; Platyhelminthes; Rhabditophora; Neodermata;
OC Trematoda; Digenea; Strigeidae; Schistosomatidae; Schistosomidae;
OC Schistosoma.
OX NCBI_Taxid=6182;
RN [1]
RP SEQUENCE FROM N.A.
RA Song X.T., Feng Z.Q., Guan X.H.;
RT "Amplification, cloning and sequence analysis of the heavy chain
RT variable region gene of monoclonal anti-idiotypic antibody NP30 of
RT Schistosoma japonicum."
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: TO IMMUNOGLOBULIN AND MAJOR HISTOCOMPATIBILITY COMPLEX
CC DOMAIN.
DR EMBL: AF282622; AAC01452.1; -.
DR InterPro: IPR003599; Ig_
DR InterPro: IPR003600; Ig_like.
DR InterPro: IPR003006; Ig_MHC.
DR InterPro: IPR003596; Ig_v.
DR Pfam: PF00047; Ig_1.
DR SMART: SM00409; Ig; 1.
DR SMART: SM00406; Ig; 1.
DR SMART: SM00410; Ig_like; 1.
DR SMART: SM00410; Ig; 1.
FT NON_TER 1
FT NON_TER 119
SQ SEQUENCE 119 AA; 13567 MW; BA93873FDF5FA6AB CRC64;

Query Match 34.6%; Score 531; DB 5; Length 119;
Best Local Similarity 80.7%; Pred. No. 3.2e-37;
Matches 96; Conservative 10; Mismatches 13; Indels 0; Gaps 0;

```

```

AC 099L25;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE SIMILAR TO RIKEN CDNA 1810060009 GENE.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Streusberg R.;
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: BC003888; AA03888.1; -.
SQ SEQUENCE 473 AA; 52449 MW; BE989B7986DA155 CRC64;

Query Match 33.2%; Score 510; DB 11; Length 473;
Best Local Similarity 45.8%; Pred. No. 1e-34;
Matches 121; Conservative 29; Mismatches 64; Indels 50; Gaps 8;

Qy 22 AOVOLQSGAELARGASVKNKSGAGYFTYTMHWKORPGGLEWIGYINPSRGYTN 81
Db 19 SOVOLQSGAELARGASVKNKSGAGYFTYTMHWKORPGGLEWIGYINPSRGYTN 78
Qy 82 YQKFKRATLTITDSSSTAYWQSLTSEDSAVVYCAR---YD-DHYSIDYMGQGT 136
Db 79 YNEKRGKATLTADKSSSTAYWQSLTSEDSAVVYCARFCSSGSIYGYGYIDYMGQGT 138
Qy 137 LTVSSAKTTPKLESEFSEARVDIVLTQSPAIMSASP-----GKRYMTCSASSVSYM 190
Db 139 LTVSSAKTTPKLESEFSEARVDIVLTQSPAIMSASP-----APSYPLAPVCGDTGSSVTLGCLVK--GYF 176
Qy 191 NMYQKSGSTSPKRWYDTSKLAGV---PAHFRGSGSGTSYSLTISGMEADATYYCQGMSSNP 247
Db 177 -----PEPTLTWNSGSLSSGVHTPEAVLQSLDLYTLSSSVTVY-----STWPSQS 222
Qy 248 WSNPFTFGSGTKLEINRADTAPT 271
Db 223 ITCNVAHPRASSTKYDKIIPRGPT 246

RESULT 4
09DBL4 PRELIMINARY; PRT; 473 AA.
AC 09DBL4;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE 1810060009RIK PROTEIN.
GN 1810060009RIK
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=C57BL/6J; TISSUE=PANCREAS;
RA MEDLINE=21085660; Pubmed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nakai I., Pesole G., Quackenbush J.,
RA Schram L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Balderelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gwinnich S., Hill D., Hofmann M., Hume D.A., Kamita M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Momberts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,

```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 5, 2002, 17:05:02 ; Search time 12.58 Seconds
(without alignments)
520.545 Million cell updates/sec

Title: US-09-424-705-2

Perfect score: 1536
Sequence: 1 MKYLLPFAAGLLLAOPA.....GSEQKLISEEDLNHHHHH 291

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

T number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/PCTUS.COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/Backlist1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	974	63.4	246	1 US-08-469-486-57	Sequence 57, Appl
2	974	63.4	246	2 US-08-469-658-57	Sequence 57, Appl
3	963.5	62.7	239	3 US-08-279-772A-8	Sequence 8, Appl
4	963.5	62.7	239	4 US-08-902-486-11	Sequence 11, Appl
5	957.5	62.3	599	1 US-08-463-163-3	Sequence 3, Appl
6	928.5	60.4	270	2 US-08-652-507-2	Sequence 2, Appl
7	919	59.8	244	2 US-08-553-497A-20	Sequence 20, Appl
8	908	59.1	246	2 US-08-553-497A-24	Sequence 24, Appl
9	905	58.9	244	2 US-08-553-497A-22	Sequence 22, Appl
10	901	58.7	242	2 US-08-553-497A-26	Sequence 26, Appl
11	898	58.5	553	2 US-08-661-052-16	Sequence 16, Appl
12	898	58.5	553	4 US-09-188-082-16	Sequence 16, Appl
13	889.5	57.9	271	2 US-08-894-922A-10	Sequence 10, Appl
14	886	57.7	242	2 US-08-553-497A-28	Sequence 28, Appl
15	874.5	56.9	273	2 US-08-403-853-18	Sequence 18, Appl
16	833.5	54.3	284	3 US-09-184-658-40	Sequence 40, Appl
17	796	51.8	223	2 US-08-190-199A-63	Sequence 63, Appl
18	768.5	51.3	252	2 US-08-894-922A-14	Sequence 14, Appl
19	788	51.3	265	2 US-08-428-257A-72	Sequence 72, Appl
20	788	51.3	269	2 US-08-491-988-3	Sequence 3, Appl
21	788	51.3	402	2 US-08-491-988-9	Sequence 9, Appl
22	788	51.3	415	2 US-08-491-988-7	Sequence 7, Appl
23	788	51.3	435	2 US-08-491-988-5	Sequence 5, Appl
24	783.5	51.0	236	2 US-08-190-199A-65	Sequence 65, Appl
25	776.5	50.6	240	2 US-08-800-198-8	Sequence 8, Appl
26	776.5	50.6	240	2 US-09-236-535-8	Sequence 8, Appl
27	773.5	50.4	222	2 US-08-190-199A-67	Sequence 67, Appl

28	771.5	50.2	240	2 US-08-956-047-25	Sequence 25, Appl
29	770.5	50.2	269	4 US-08-646-265A-109	Sequence 109, App
30	761	49.5	235	2 US-08-190-199A-61	Sequence 61, Appl
31	752	49.0	553	2 US-08-263-911-7	Sequence 7, Appl
32	750	48.8	282	2 US-08-860-174A-10	Sequence 10, Appl
33	738.5	48.1	284	4 US-08-564-164A-2	Sequence 2, Appl
34	737	48.0	269	4 US-09-070-408-132	Sequence 132, App
35	724	47.1	249	2 US-08-797-689-18	Sequence 18, Appl
36	723.5	47.1	265	2 US-08-403-853-16	Sequence 16, Appl
37	720.5	46.9	553	2 US-08-263-911-9	Sequence 9, Appl
38	715.5	46.6	483	2 US-08-392-338A-19	Sequence 19, Appl
39	715.5	46.6	483	3 US-09-166-750-19	Sequence 19, Appl
40	715.5	46.6	483	3 US-09-166-093-19	Sequence 19, Appl
41	715.5	46.6	483	3 US-09-172-019-19	Sequence 19, Appl
42	715.5	46.6	483	3 US-09-166-094-19	Sequence 19, Appl
43	712.5	46.4	241	1 US-08-235-838-11	Sequence 11, Appl
44	712.5	46.4	241	2 US-08-465-473B-11	Sequence 11, Appl
45	712.5	46.4	637	1 US-08-235-838-16	Sequence 16, Appl

ALIGNMENTS

RESULT 1
US-08-469-486-57
Sequence 57, Application US/08469486
Patent No. 5739281
GENERAL INFORMATION:
APPLICANT: Thoegeisen, Hans Christian
APPLICANT: Holteit, Thor Las
APPLICANT: Etzerodt, Michael
TITLE OF INVENTION: Improved method for the refolding of
NUMBER OF SEQUENCES: 58
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson
STREET: 225 Franklin Street
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version
SOFTWARE: #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/469,486
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/192,060
FILING DATE: February 4, 1994
ATTORNEY/AGENT INFORMATION:
NAME: Paul T. Clark
REGISTRATION NUMBER: 30,162
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617 542 5070
TELEFAX: 617 542 8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 57:
SEQUENCE CHARACTERISTICS:
LENGTH: 246 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-469-486-57

Query Match 63.4%; Score 974; DB 1; Length 246;

GenCore version 4.5
Copyright (c) 1993 - 2000 Compen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 5, 2002, 18:23:18 ; Search time 168.12 Seconds
(without alignments)
122.388 Million cell updates/sec

Title: US-09-424-705-7

Perfect score: 24
Sequence: 1 gtatgaagcgtgatgcatc 24

Scoring table: IDENTITY NUC
Gapop 10.0, Gapept 1.0

Searched: 930621 seqs, 428662619 residues

number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

N.Geneseq_1101.*
1: /SIDSL/gcgcdata/geneseq/geneseqn/NA1980.DAT.*
2: /SIDSL/gcgcdata/geneseq/geneseqn/NA1981.DAT.*
3: /SIDSL/gcgcdata/geneseq/geneseqn/NA1982.DAT.*
4: /SIDSL/gcgcdata/geneseq/geneseqn/NA1983.DAT.*
5: /SIDSL/gcgcdata/geneseq/geneseqn/NA1984.DAT.*
6: /SIDSL/gcgcdata/geneseq/geneseqn/NA1985.DAT.*
7: /SIDSL/gcgcdata/geneseq/geneseqn/NA1986.DAT.*
8: /SIDSL/gcgcdata/geneseq/geneseqn/NA1987.DAT.*
9: /SIDSL/gcgcdata/geneseq/geneseqn/NA1988.DAT.*
10: /SIDSL/gcgcdata/geneseq/geneseqn/NA1989.DAT.*
11: /SIDSL/gcgcdata/geneseq/geneseqn/NA1990.DAT.*
12: /SIDSL/gcgcdata/geneseq/geneseqn/NA1991.DAT.*
13: /SIDSL/gcgcdata/geneseq/geneseqn/NA1992.DAT.*
14: /SIDSL/gcgcdata/geneseq/geneseqn/NA1993.DAT.*
15: /SIDSL/gcgcdata/geneseq/geneseqn/NA1994.DAT.*
16: /SIDSL/gcgcdata/geneseq/geneseqn/NA1995.DAT.*
17: /SIDSL/gcgcdata/geneseq/geneseqn/NA1996.DAT.*
18: /SIDSL/gcgcdata/geneseq/geneseqn/NA1997.DAT.*
19: /SIDSL/gcgcdata/geneseq/geneseqn/NA1998.DAT.*
20: /SIDSL/gcgcdata/geneseq/geneseqn/NA1999.DAT.*
21: /SIDSL/gcgcdata/geneseq/geneseqn/NA2000.DAT.*
22: /SIDSL/gcgcdata/geneseq/geneseqn/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	24	100.0	24	AAV73336	Mouse OKT3 variant
2	24	100.0	898	AAV73335	Mouse OKT3 variant
3	24	100.0	1653	AAZ43432	Fv-antibody constr
4	24	100.0	1698	AAZ43431	Fv-antibody constr
5	24	100.0	1794	AAV73337	Mouse bispecific a
6	22.4	93.3	1570	AAO12637	Monoclonal antibody
7	22.4	93.3	1611	AAZ30332	Nucleotide sequen
8	17.6	73.3	47	AAZ65503	Human map-related
9	17.6	73.3	587	AAF08479	Fusarium venenatum
10	17.4	72.5	249	AAI21165	Probe #11098 for g
11	17.4	72.5	249	AAI46432	Probe #15118 used

12	17.4	72.5	249	22	AAI06879	Probe #6870 used t
13	17.4	72.5	423	20	AAZ40429	Human RAD1 5' EST
14	17.4	72.5	474	22	AAI11958	Probe #1891 for ge
15	17.4	72.5	474	22	AAI33289	Probe #1975 used t
16	17.4	72.5	474	22	AAI01890	Probe #1881 used t
17	17.4	72.5	551	20	AAZ40425	Human RAD1 5' EST
18	17.4	72.5	846	20	AAZ40423	Human RAD1 gene.
19	17.4	72.5	849	20	AAZ78787	Human RAD1 clone 4
20	17.4	72.5	1145	20	AAZ40434	Human RAD1 clone 6
21	17.4	72.5	1240	20	AAZ40436	Human RAD1 clone 4
22	17.4	72.5	1396	22	AAI15084	Human CDNA sequenc
23	17.2	71.7	353	21	AAO2746	Human secreted pro
24	17.2	71.7	370	21	AAZ82128	N. meningitidis pa
25	17.2	71.7	846	21	AAZ30869	Streptococcus equi
26	16.8	70.0	1898	22	AAZ6854	Human secreted pro
27	16.8	70.0	1919	21	AAZ77751	Human cancer assoc
28	16.8	70.0	1973	22	AAI14117	Human CDNA sequenc
29	16.8	70.0	2560	19	AAV52247	Streptococcus pneu
30	16.8	70.0	3805	20	AAZ00048	HGF2 gene. Homo
31	16.6	69.2	386	22	AAZ64409	Novel human polynu
32	16.6	69.2	605	22	AAH31723	Human olfactory re
33	16.6	69.2	764	21	AAO1473	Human secreted pro
34	16.6	69.2	945	22	AAH31828	Human olfactory re
35	16.6	69.2	1807	19	AAV43719	Cancer associated
36	16.6	69.2	1904	21	AAZ78093	Human cancer assoc
37	16.6	69.2	2290	12	AAI15170	Bacillus thuringie
38	16.6	69.2	2297	20	AAZ22303	CDNA encoding a mo
39	16.6	69.2	6675	20	AAZ84352	Stealth virus nucl
40	16.2	67.5	531	21	AAZ46600	Zea mays DNA fragm
41	16.2	67.5	700	22	AAH93253	Human inflammatory
42	16.2	67.5	756	22	AAH65556	C glutamic codin
43	16.2	67.5	952	21	AAH35913	Arabidopsis thalia
44	16.2	67.5	1258	22	AAI60912	Human polynucleoti
45	16.2	67.5	9819	21	AAZ94999	Cancer specific ge

ALIGNMENTS

RESULT 1	AAV73336	standard; DNA: 24 BP.
ID	AAV73336	
AC	AAV73336	
DT	26-FEB-1999	(first entry)
DE	Mouse OKT3 variant antibody primer SK1 DNA.	
XX	OKT3; monoclonal antibody; Mab; point mutation; transplant rejection;	
KW	organ recipient; diagnosis; tumour; therapy; primer; ss.	
OS	Synthetic.	
OS	Mus sp.	
PN	DE19721700-C1.	
PD	19-NOV-1998.	
XX		
XX	23-MAY-1997;	97DE-1021700.
XX		
XX	23-MAY-1997;	97DE-1021700.
XX		
PA	(DEKR-) DEUT KREBSFORSCHUNGSENTRUM.	
XX	Kipriyanov S, Little M, Moldenhauer G;	
XX	WPI; 1998-596150/51.	
PT	Monoclonal antibody OKT3 with point-mutation - where cysteine is	
PT	replaced by another polar amino acid, useful for controlling	
PT	transplant rejection, and in tumour diagnostics and therapy	
XX		

PS Claim 7; Column 3; 8pp; German.

CC This sequence is a primer used in the construction of a monoclonal
 CC antibody (Mab) derived from OKT3 with a point-mutation where Cys at
 CC position H100A is replaced with another polar amino acid, in this example
 CC Ser. The Mab is used in lowering or eliminating the transplant rejection
 CC in an organ recipient and for diagnostic methods for tumours and tumour
 CC therapy.

XX
 SQ Sequence 24 BP; 7 A; 4 C; 6 G; 7 T; 0 other;

Query Match 100.0%; Score 24; DB 19; Length 24;

Best Local Similarity 100.0%; Pred. No. 0.019;
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gtagtcaaggctgtaatgcatc 24
 1 gtagtcaaggctgtaatgcatc 24

RESULT 2

AAV73335/c

ID AAV73335 standard; DNA; 898 BP.

XX AAV73335;

XX 26-FEB-1999 (first entry)

XX Mouse OKT3 variant antibody DNA.

XX OKT3; monoclonal antibody; Mab; point mutation; transplant rejection;

KM organ recipient; diagnosis; tumour; therapy; ss.

XX Mus sp.

OS Synthetic.

XX DE19721700-cl.

XX 19-NOV-1998.

XX 23-MAY-1997; 97DE-1021700.

XX 23-MAY-1997; 97DE-1021700.

XX (DEKR-) DEUT KREBSFORSCHUNGSZENTRUM.

XX Kipriyanov S, Little M, Moldenhauer G;

DR WPI: 1998-596150/51.

DR P-PSDB; AAW82316.

PT Monoclonal antibody OKT3 with point-mutation - where cysteine is
 PT replaced by another polar amino acid, useful for controlling
 PT transplant rejection, and in tumour diagnostics and therapy

PS Claim 3; Fig 2; 8pp; German.

XX This sequence encodes a monoclonal antibody (Mab) derived from OKT3 with
 CC a point-mutation where Cys at position H100A is replaced with another
 CC polar amino acid, in this example Ser. The Mab is used in lowering or
 CC eliminating the transplant rejection in an organ recipient, and for
 CC diagnostic methods for tumours and tumour therapy.

XX Sequence 898 BP; 245 A; 236 C; 217 G; 200 T; 0 other;

Query Match 100.0%; Score 24; DB 19; Length 898;

Best Local Similarity 100.0%; Pred. No. 0.036;
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gtagtcaaggctgtaatgcatc 24
 1 gtagtcaaggctgtaatgcatc 24

Db 417 GTAGTCAAGCTGTAATGATCATC 394

RESULT 3

AAZ43432/c

ID AAZ43432 standard; DNA; 1653 BP.

XX AAZ43432;

XX 18-FEB-2000 (first entry)

XX Fv-antibody construct containing antibody 9E10 epitope DNA.

XX Multivalent; Fv-antibody; antiviral; antibacterial; cytostatic;

XX diagnosis; therapy; disease; ss.

XX Synthetic.

XX DE19819846-A1.

XX 11-NOV-1999.

XX 05-MAY-1998; 98DE-1019846.

XX 05-MAY-1998; 98DE-1019846.

XX (DEKR-) DEUT KREBSFORSCHUNGSZENTRUM.

XX Little M, Kipriyanov S;

XX WPI: 2000-024472/03.

XX Multivalent Fv-antibody constructs with at least four variable domains

XX connected by 1, 2 and 3 peptide linkers

XX Example 1; Fig 6; 14pp; German.

XX This invention describes a novel multivalent Fv-antibody construct with
 CC at least four variable domains that are connected to one another by 1,
 CC 2 and 3 peptide linkers. The construct has antiviral, antibacterial
 CC and cytostatic activity. The multivalent Fv-antibody constructs are
 CC useful for the diagnosis and/or therapy of disease, especially viral,
 CC bacterial or tumor diseases. The multivalent Fv-antibody constructs have
 CC increased stability when in the form of a single chain dimer. This
 CC sequence encodes a tetraivalent Fv antibody construct composed of the
 CC antibody 9E10 epitope in expression plasmid pDSC319-SL

XX Sequence 1653 BP; 432 A; 429 C; 424 G; 368 T; 0 other;

Query Match 100.0%; Score 24; DB 21; Length 1653;

Best Local Similarity 100.0%; Pred. No. 0.041;
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gtagtcaaggctgtaatgcatc 24
 417 GTAGTCAAGCTGTAATGATCATC 394

RESULT 4

AAZ43431/c

ID AAZ43431 standard; DNA; 1698 BP.

XX AAZ43431;

XX 18-FEB-2000 (first entry)

XX Fv-antibody construct containing antibody 9E10 epitope DNA.

XX Multivalent; Fv-antibody; antiviral; antibacterial; cytostatic;

XX diagnosis; therapy; disease; ss.

XX Synthetic.

XX DE19819846-A1.
PN 11-NOV-1999.
XX
XX 05-MAY-1998; 98DE-1019846.
XX
XX 05-MAY-1998; 98DE-1019846.
XX
XX (DEKR-) DEUT KREBSFORSCHUNGSZENTRUM.
XX
XX Little M, Kipriyanov S;
XX WPI; 2000-024472/03.
XX P-PSDB; AAV50822.
XX
XX Multivalent Fv-antibody constructs with at least four variable domains
XX connected by 1, 2 and 3 peptide linkers
XX
XX Example 1; Fig 5; 14pp; German.
XX
XX This invention describes a novel multivalent Fv-antibody construct with
XX at least four variable domains that are connected to one another by 1,
XX 2 and 3 peptide linkers. The construct has antiviral, antibacterial
XX and cytostatic activity. The multivalent Fv-antibody constructs are
XX useful for the diagnosis and/or therapy of disease, especially viral,
XX bacterial or tumor diseases. The multivalent Fv-antibody constructs have
XX increased stability when in the form of a single chain dimer. This
XX sequence encodes a bivalent Fv antibody construct composed of the
XX antibody 9E10 epitope in expression plasmid pDISC3X19-LT
XX
XX Sequence 1698 BP; 433 A; 434 C; 450 G; 381 T; 0 other;
XX

Query Match 100.0%; Score 24; DB 21; Length 1698;
Best Local Similarity 100.0%; Pred. No. 0.041;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 gtagtcaggctgtaatgcatc 24
DB 417 GTAGTCAGGCTGTATGATCATC 394
|||||

RESULT 5
AAV73337/C
ID AAV73337 standard; DNA; 1794 BP.
XX
XX AAV73337;
XX
XX 26-FEB-1999 (first entry)
XX
DE Mouse bispecific antibody construct variant OKT3/anti-CD19 DNA.
XX
XX OKT3; monoclonal antibody; Mab; point mutation; transplant rejection;
XX organ recipient; diagnosis; tumour; therapy; diabody; anti-CD19;
XX anti-CD3; ss.
XX
XX Mus sp.
XX Synthetic.
XX
XX
XX Key Location/Qualifiers
XX CDS 28..894
XX FT /*tag= a
XX FT /*product= "mutant OKT3 with anti-CD19 insert"
XX FT 922..1788
XX FT /*tag= b
XX FT /*product= "mutant OKT3 with anti-CD3 insert"
XX
XX DE19721700-C1.
XX
XX 19-NOV-1998.
XX
XX 23-MAY-1997; 97DE-1021700.
XX

XX 23-MAY-1997; 97DE-1021700.
XX
XX (DEKR-) DEUT KREBSFORSCHUNGSZENTRUM.
XX
XX Kipriyanov S, Little M, Moldenhauer G;
XX WPI; 1998-596150/51.
XX P-PSDB; AAW82317, AAW82482.
XX
XX Monoclonal antibody OKT3 with point-mutation - where cysteine is
XX replaced by another polar amino acid, useful for controlling
XX transplant rejection, and in tumour diagnostics and therapy
XX
XX Disclosure; Fig 3; 8pp; German.
XX
XX This sequence encodes a monoclonal antibody (Mab) diabody derived from
XX OKT3 with a point-mutation where Cys at position H100A is replaced with
XX another polar amino acid, in this example Ser. The diabody encodes two
XX OKT3 proteins, one which has anti-CD19 activity and one with anti-CD3
XX activity. The Mab is used in lowering or eliminating the transplant
XX rejection in an organ recipient and for diagnostic methods for tumours
XX and tumour therapy.
XX
XX Sequence 1794 BP; 480 A; 468 C; 450 G; 396 T; 0 other;
XX

Query Match 100.0%; Score 24; DB 19; Length 1794;
Best Local Similarity 100.0%; Pred. No. 0.041;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 gtagtcaggctgtaatgcatc 24
DB 417 GTAGTCAGGCTGTATGATCATC 394
|||||

RESULT 6
AAQ12637/C
ID AAQ12637 standard; DNA; 1570 BP.
XX
XX AAQ12637;
XX
XX 03-OCT-1991 (first entry)
XX
XX Monoclonal antibody OK3T heavy chain coding sequence.
XX
XX OK3T; heavy chain; humanised antibodies; CDR-grafting; ss.
XX
XX Mus musculus.
XX
XX
XX Key Location/Qualifiers
XX FH sig_peptide 41..97
XX FT /*tag= a
XX FT mat_peptide 98..1447
XX FT /*tag= b
XX FT /*product= OK3T heavy chain
XX
XX WO9109967-A.
XX PN WO9109968-A.
XX PN
XX 11-JUL-1991.
XX
XX 21-DEC-1990; 90WO-GB02017.
XX PF
XX 21-DEC-1990; 90WO-GB02017.
XX PR
XX 21-DEC-1989; 89GB-0028874.
XX
XX (CELL-) CELTECH LTD.
XX
XX Adair JR, Athwal DS, Emtage JS;
XX WPI; 1991-222915/30.
XX P-PSDB; AAR13061.
XX
XX

```

XX New humanised antibodies comprising CDR grafted antibody - with
PT heavy and light chains, for use in in vivo therapy and diagnosis
XX
XX Disclosure; Fig 2a; 91pp; English.
PS
CC The OK3T heavy chain sequence was isolated from a cDNA library
CC prepared from OK3T producing cells. The library was screened with a
CC probe complementary to a sequence in the mouse IgG2a constant CH1
CC domain region. The OK3T sequence was used in CDR-grafting experiments
CC to prepare humanised antibodies.
XX
SQ Sequence 1570 BP; 447 A; 443 C; 356 G; 324 T; 0 other;

Query Match          93.3%; Score 22.4; DB 12; Length 1570;
Best Local Similarity 95.8%; Pred. No. 0.23;
Matches 23; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

1 gtatcaagcgtgaatgcatc 24
|||||
Db 421 GTAGTCACGCGCATATGATCATC 398

RESULT 7
AAZ30332/C
ID AAZ30332 standard; cDNA; 1611 BP.
XX
AC AAZ30332;
XX
DT 11-FEB-2000 (first entry)
XX
DE Nucleotide sequence of the bscCD19xCD3 antibody.
XX
KW bscCD19xCD3 antibody; bispecific single-chain fragment; CD19 antigen;
KW CD3 antigen; CD19-positive target cell; T-cell stimulation;
KW cytotoxic T-lymphocyte; B-cell malignancy; myasthenia gravis;
KW B-cell mediated autoimmune disease; Morbus Basedow;
KW Hashimoto thyroiditis; Goodpasture syndrome; B-cell depletion;
KW non-Hodgkin lymphoma; gene therapy; cancer; viral disease; ds.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT CDS 11..1606
FT /tag= a
FT /product= "bscCD19xCD3 antibody"
FT 11..67
FT /tag= b
FT 92..424
FT /tag= c
FT /note= "encode VL-CD19"
FT 470..841
FT /tag= d
FT /note= "encode VH-CD19"
FT 857..1213
FT /tag= e
FT /note= "encodes VH-CD3"
FT 1268..1585
FT /tag= f
FT /note= "encodes VL-CD3"
FT 1586..1603
FT /tag= g
FT /note= "encodes a His tag"
XX
PN WO954440-A1.
XX
PD 28-OCT-1999.
XX
PF 21-APR-1999; 99WO-EP02693.
XX
PR 21-APR-1998; 98EP-0107269.
XX

```

```

PA (DOER/) DOERKEN B.
PA (RIET/) RIETHMUELLER G.
XX
PI Kufer P, Lutterbuese R, Bargou R, Loeffler A;
XX
DR WPI; 2000-013241/01.
DR P-PSDB; AAY43749.
XX
PT Novel multifunctional polypeptide for treating B-cell malignancies
PT especially non-Hodgkin lymphoma
XX
PS Example 2; Fig 8; 91pp; English.
XX
CC The present sequence encodes a bscCD19xCD3 antibody. This antibody
CC is a bispecific single-chain polypeptide comprising domains providing
CC binding-site of immunoglobulin chains or antibodies specifically
CC recognizing CD19 and CD3 antigen. The polypeptide destroys CD19-positive
CC target cells without any need of T-cell pre and/or co-stimulation, by
CC recruiting cytotoxic T-lymphocytes and so specific lysis by T-cells
CC rather than a direct effect by an antibody is achieved. The bispecific
CC single-chain polypeptides, or nucleotides encoding them, are used for
CC the treatment of B-cell malignancies, B-cell mediated autoimmune
CC diseases like myasthenia gravis, Morbus Basedow, Hashimoto thyroiditis
CC or Goodpasture syndrome or for the depletion of B-cells and more
CC particularly non-Hodgkin lymphoma in mammals preferably human. They can
CC also delay the pathological conditions caused by these diseases, and
CC can be used for detecting these diseases. The polynucleotide is used
CC for gene therapy. The polypeptides are also used for identifying
CC compounds modulating B-cell/T-cell mediated immune response with can in
CC turn be used for treating cancer, its related diseases and also for
CC inhibiting viral diseases by preventing viral infection.
XX
SQ Sequence 1611 BP; 402 A; 396 C; 440 G; 373 T; 0 other;

Query Match          93.3%; Score 22.4; DB 21; Length 1611;
Best Local Similarity 95.8%; Pred. No. 0.23;
Matches 23; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

1 gtatcaagcgtgaatgcatc 24
|||||
Db 1180 GTAGTCACGCGCATATGATCATC 1157

RESULT 8
AAZ66503/C
ID AAZ66503 standard; DNA; 47 BP.
XX
AC AAZ66503;
XX
DT 10-SEP-2001 (first entry)
XX
DE Human map-related diallelic marker SEQ ID NO:850.
XX
KW Human genome; diallelic marker; high density disequilibrium map;
KW genomic map; haplotype; phenotype; polymorphic base; genotyping;
KW haplotyping; hybridisation; identification; characterisation;
KW diagnosis; single nucleotide polymorphism; SNP; ds.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT variation replace(24,C)
FT /tag= a
FT /standard_name= "single nucleotide polymorphism"
XX
PN WO954500-A2.
XX
PD 28-OCT-1999.
XX
PF 21-APR-1999; 99WO-IB00822.
XX
PR 21-APR-1998; 98US-0082614.
XX

```


PR 23-NOV-1998; 980US-0109732.
 XX
 PA (GEST) GENSET.
 XX
 PI Cohen D, Blumenfeld M, Chumakov I;
 XX
 DR WPI: 2000-013267/01.
 XX
 PS Novel biallelic markers used to construct a high density disequilibrium
 PT map of the human genome
 PR
 PS Claim 1: Page 412: 2745pp: English.
 XX
 CC AA265654 to AA269578 represent human biallelic markers from the present
 CC invention, which contain a polymorphic base at position 24 of their
 CC nucleotide sequences. AA269579 to AA277440 represent amplification
 CC primers for the biallelic markers. The biallelic markers of the
 CC invention have a variety of uses: they can be used for high density
 CC mapping of the human genome, and in complex association studies and
 CC haplotyping studies which are useful in determining the genetic basis
 CC for disease states. Compositions and methods of the invention can also
 CC be useful for the identification of the targets for the development of
 CC pharmaceutical agents and diagnostic methods, as well as the
 CC characterisation of the differential efficacious responses to and side
 CC effects from pharmaceutical agents acting on a disease as well as other
 CC treatment.
 CC N.B. The SEQ ID Nos 2852, 2913, 2974, 3035, 3096, 3157, 3227, 3297
 CC and 3367, are not actually given a sequence in the Sequence Listing
 CC from the present invention.
 CC
 XX
 SQ Sequence 47 BP; 15 A; 11 C; 5 G; 16 T; 0 other;

Query Match 73.3%; Score 17.6; DB 21: Length 47;
 Best Local Similarity 83.3%; Pred. No. 24;
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 gtagtcaagcgtgtaatgcatc 24
 ||||| | ||||| ||||| |||||
 DB 37 GTAGTACACGCGTATATGATGATC 14

RESULT 9
 AAF08479
 ID AAF08479 standard; cDNA; 587 BP.
 XX
 AC AAF08479;
 XX
 DT 13-MAR-2001 (first entry)
 XX
 DE Fusarium venenatum EST SEQ ID NO:1002.
 XX
 KW Multiple gene expression; filamentous fungal cell; EST;
 KW expressed sequence tag; Fusarium venenatum; Aspergillus niger;
 KW Aspergillus oryzae; Trichoderma reesei; identification; recombination;
 KW culture condition; environmental stress; spore morphogenesis;
 KW metabolic pathway engineering; catabolic pathway engineering; ss.
 XX
 OS Fusarium venenatum.
 XX
 PN WO200056762-A2.
 XX
 PD 28-SEP-2000.
 XX
 PE 22-MAR-2000; 2000WO-US07781.
 XX
 PR 22-MAR-1999; 99US-0273623.
 XX
 PA (NOVO) NOVO NORDISK BIOTECH INC.
 PA (NOVO) NOVO NORDISK AS.
 XX
 PI Betka RM, Rey MM, Shuster JR, Kauppinen S, Clausen IG, Olsen PB;
 XX

DR WPI: 2000-594572/56.
 XX
 XX
 PR Monitoring differential expression of genes in filamentous fungal cells
 PT uses fluorescence-labeled nucleic acids isolated from the cells and a
 PT substrate of expressed sequence tags -
 PS
 PS Claim 86: Page 764-765; 3161pp: English.
 XX
 CC The present invention describes a method for monitoring differential
 CC expression of genes in a first filamentous fungal (FF) cell relative to
 CC expression of the same genes in one or more second filamentous fungal
 CC cells. The method uses fluorescence-labeled nucleic acids isolated from
 CC the FF cells and a substrate of expressed sequence tags (EST). The ESTs
 CC are used in the methods for monitoring differential expression of genes
 CC in a first filamentous fungal (FF) cell relative to expression of the
 CC same genes in one or more second filamentous fungal cells. Monitoring
 CC the global expression of genes from FF cells allows the production
 CC potential of the microorganisms to be improved. New genes may be
 CC discovered, possible functions of unknown open reading frames can be
 CC identified and gene copy number variation and stability can be
 CC monitored. The expression of genes can be used to study how FF cells
 CC adapt to changes in culture conditions, environmental stress, spore
 CC morphogenesis, recombination, metabolic or catabolic pathway
 CC engineering. Using ESTs provides several advantages over genomic or
 CC random cDNA clones including elimination of redundancy as one spot on an
 CC array equals one gene or open reading frame, and organisation of the
 CC microarrays based on function of the gene products to facilitate
 CC analysis of the results. AAF07478 to AAF11247 represents ESTs from
 CC Fusarium venenatum; AAF11248 to AAF11853 represents ESTs from
 CC niger; AAF11854 to AAF14878 represents ESTs from Aspergillus oryzae; and
 CC AAF14879 to AAF1537 represents ESTs from Trichoderma reesei, which are
 CC all specifically claimed in the present invention.
 CC
 XX
 SQ Sequence 587 BP; 146 A; 133 C; 131 G; 172 T; 5 other;

Query Match 73.3%; Score 17.6; DB 21: Length 587;
 Best Local Similarity 83.3%; Pred. No. 38;
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 gtagtcaagcgtgtaatgcatc 24
 ||||| | ||||| ||||| |||||
 DB 443 gtgtcttctgctgtaatgcatc 466

RESULT 10
 AAT21165
 ID AAT21165 standard; DNA; 249 BP.
 XX
 AC AAT21165;
 XX
 DT 12-OCT-2001 (first entry)
 XX
 DE Probe #11098 for gene expression analysis in human cervical cell sample.
 KW Probe; human; microarray; gene expression; cervical epithelial cell;
 KW cervical cancer; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200157278-A2.
 XX
 PD 09-AUG-2001.
 XX
 PE 30-JAN-2001; 2001WO-US00670.
 XX
 PR 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 PR

XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR:
XX
DR WPI; 2001-488901/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human cervical epithelial cells -
XX
PS Claim 25: SEQ ID No 11098; 487pp; English.
XX
CC The present invention relates to human single exon nucleic acid probes
CC (SENPs). The present sequence is one such probe. The SENPs are derived
CC from human HeLa cells. The SENPs can be used to produce a single exon
CC microarray, which can be used for measuring human gene expression in a
CC sample derived from human cervical epithelial cells. By measuring gene
CC expression, the probes are therefore useful in grading and/or staging
CC of diseases of the cervix, notably cervical cancer.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 249 BP; 58 A; 59 C; 60 G; 72 T; 0 other;
XX
Query Match 72.5%; Score 17.4; DB 22; Length 249;
Best Local Similarity 94.7%; Pred. No. 41;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX
QY 6 caagctgtaatgcatc 24
Db 147 caagctgtaatgcatc 165
XX
RESULT 11
AA16432
ID AA16432 standard; DNA; 249 BP.
XX
AC AA16432;
XX
DT 17-OCT-2001 (first entry)
XX
DE Probe #15118 used to measure gene expression in human placenta sample.
XX
KW Probe: microarray; human; placenta; antenatal diagnosis;
KW genetic disorder; ss.
XX
PA Homo sapiens.
XX
PN WO200157272-A2.
XX
PD 09-AUG-2001.
XX
PE 30-JAN-2001; 2001WO-US00663.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR:
XX
DR WPI; 2001-48897/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human placenta -
XX

PS Claim 25: SEQ ID No 15118; 654pp; English.
XX
CC The present invention relates to single exon nucleic acid probes (SENPs).
CC The present sequence is one such probe. The probes are useful for
CC producing a microarray for predicting, measuring and displaying gene
CC expression in samples derived from human placenta. The probes are useful
CC for antenatal diagnosis of human genetic disorders.
XX
SQ Sequence 249 BP; 58 A; 59 C; 60 G; 72 T; 0 other;
XX
Query Match 72.58%; Score 17.4; DB 22; Length 249;
Best Local Similarity 94.7%; Pred. No. 41;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX
QY 6 caagctgtaatgcatc 24
Db 147 caagctgtaatgcatc 165
XX
RESULT 12
AA106879
ID AA106879 standard; DNA; 249 BP.
XX
AC AA106879;
XX
DT 09-OCT-2001 (first entry)
XX
DE Probe #6870 used to measure gene expression in human breast sample.
XX
KW Probe: human; breast disease; breast cancer; development disorder; ss;
KW inflammatory disease; proliferative breast disease; non-carcinoma tumour.
XX
OS Homo sapiens.
XX
PN WO200157270-A2.
XX
PD 09-AUG-2001.
XX
PE 29-JAN-2001; 2001WO-US00661.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR:
XX
DR WPI; 2001-476286/51.
XX
PT Novel single exon nucleic acid probe used to measuring gene expression
PT in a human breast -
XX
PS Claim 25: SEQ ID No 6870; 322pp; English.
XX
CC The present invention relates to novel single exon nucleic acid probes.
CC The present sequence is one such probe. The probes are useful for
CC measuring human gene expression in a human breast sample, where the probe
CC hybridizes at high stringency to a nucleic acid expressed in the human
CC breast. The probes are useful for predicting, diagnosing, grading,
CC staging, monitoring and prognosing diseases of the human breast,
CC particularly those diseases with polygenic aetiology. The diseases
CC include: breast cancer, disorders of development, inflammatory diseases
CC of the breast, fibrocystic changes, proliferative breast disease and
CC non-carcinoma tumours.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 249 BP; 58 A; 59 C; 60 G; 72 T; 0 other;

Query Match

Best Local Similarity 72.5%; Score 17.4; DB 22; Length 249;
Pred. No. 41;

Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 6 caagctgtatgatcatc 24

DB 147 caagctgtatgatcatc 165

RESULT 13

AAZ40429/C

ID AAZ40429 standard; DNA: 423 BP.

XX AAZ40429;

DT 15-FEB-2000 (first entry)

XX Human RAD1 5' EST AAZ27739.

XX Human; RAD1: expressed sequence tag; EST; Schizosaccharomyces pombe;
cell cycle checkpoint; genetic alteration; locus; disease; cancer;
bladder; head and neck; small cell lung tumour; immune disorder; ss;
proliferative disorder; cirrhosis; rheumatoid arthritis; modulator.

OS Homo sapiens.

PN WO9949042-A1.

PD 30-SEP-1999.

PF 29-MAR-1999; 99WO-US06714.

PR 27-MAR-1998; 98US-0049792.

PA (ICOS-) ICOS CORP.

PI Herendeen DR;

DR WPI; 1999-572202/48.

XX New isolated human Rad1 gene, used to develop products for treating
PT e.g. cancers, immune and proliferative disorders, cirrhosis and
PT rheumatoid arthritis -

XX Example 1; Page 56; 80pp; English.

XX Sequences AAZ40425-240430 represent expressed sequence tags (ESTs)
CC spanning the coding region for a human RAD1 protein (AAZ40423). The EST
CC were isolated by a search of an EST database using the
CC Schizosaccharomyces pombe Rad1 sequence as the query sequence. Rad1
CC polypeptides are involved in cell cycle checkpoints. The polynucleotides
CC can be used for identifying genetic alterations in a Rad1 locus that
CC underlies disease states including cancer (e.g. bladder, head and neck
CC cancers as well as small cell lung tumours), immune and proliferative
CC disorders, cirrhosis, and rheumatoid arthritis. The Rad1 polypeptides can
CC be used for identifying modulators of Rad1 activities. Such modulators
CC can be used for treating disease states involving Rad1.

SO Sequence 423 BP; 102 A; 107 C; 111 G; 103 T; 0 other;

Query Match

Best Local Similarity 72.5%; Score 17.4; DB 20; Length 423;
Pred. No. 45;

Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 6 caagctgtatgatcatc 24

DB 227 CAAGCTGTACTGATCATC 209

RESULT 14

AAI1958

XX AAI1958 standard; DNA: 474 BP.

DT 12-OCT-2001 (first entry)

XX Probe #1891 for gene expression analysis in human cervical cell sample.

XX Probe; human; microarray; gene expression; cervical epithelial cell;

XX cervical cancer; ss.

OS Homo sapiens.

PN WO200157278-A2.

PD 09-AUG-2001.

PF 30-JAN-2001; 2001WO-US00670.

PR 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

PA (MOLE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DR;

DR WPI; 2001-488901/53.

PT Human genome-derived single exon nucleic acid probes useful for

PT analyzing gene expression in human cervical epithelial cells -

XX Claim 25; SEQ ID No 1891; 487pp; English.

XX The present invention relates to human single exon nucleic acid probes

XX (SENP). The present sequence is one such probe. The SENPs are derived

XX from human HeLa cells. The SENPs can be used to produce a single exon

XX microarray, which can be used for measuring human gene expression in a

XX sample derived from human cervical epithelial cells. By measuring gene

XX expression, the probes are therefore useful in grading and/or staging

XX of diseases of the cervix, notably cervical cancer.

XX Note: The sequence data for this patent did not form part of the printed

XX CC specification, but was obtained in electronic format directly from WIPO

XX at ftp.wipo.int/pub/published_pct_sequences.

SO Sequence 474 BP; 136 A; 91 C; 94 G; 153 T; 0 other;

Query Match

Best Local Similarity 72.5%; Score 17.4; DB 22; Length 474;
Pred. No. 46;

Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 6 caagctgtatgatcatc 24

DB 371 caagctgtatgatcatc 389

RESULT 15

AAI33289

XX AAI33289 standard; DNA: 474 BP.

DT 17-OCT-2001 (first entry)

XX Probe #1975 used to measure gene expression in human placenta sample.

XX Probe: microarray; human; placenta; antenatal diagnosis;
 KW genetic disorder; ss.
 XX Homo sapiens.

XX WO200157272-A2.
 PN

XX 09-AUG-2001.
 PD

XX 30-JAN-2001; 2001WO-US00663.
 PF

XX 04-FEB-2000; 2000US-0180312.
 PR

XX 26-MAY-2000; 2000US-0207456.
 PR

XX 30-JUN-2000; 2000US-0608408.
 PR

XX 03-AUG-2000; 2000US-0632366.
 PR

XX 21-SEP-2000; 2000US-0234687.
 PR

XX 27-SEP-2000; 2000US-0236359.
 PR

XX 04-OCT-2000; 2000GB-0024263.
 PR

XX (MOLE-) MOLECULAR DYNAMICS INC.
 PI

XX Penn SG, Hanzel DK, Chen W, Rank DR;
 PI

XX WPI; 2001-488897/53.
 DR

XX Human genome-derived single exon nucleic acid probes useful for
 PT analyzing gene expression in human placenta -

XX Claim 25; SEQ ID No 1975; 654bp; English.
 PS

XX The present invention relates to single exon nucleic acid probes (SENP).
 CC The present sequence is one such probe. The probes are useful for

XX producing a microarray for predicting, measuring and displaying gene
 CC expression in samples derived from human placenta. The probes are useful

XX for antenatal diagnosis of human genetic disorders.
 CC

XX Sequence 474 BP; 136 A; 91 C; 94 G; 153 T; 0 other;
 SQ

Query Match 72.5%; Score 17.4; DB 22; Length 474;
 Best Local Similarity 94.7%; Pred. No. 46;
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 OY 6 caagcgtatgcatc 24
 |||||
 Db 371 caagcgtatgcatc 389

Search completed: March 5, 2002, 19:19:05
 Job time: 3347 sec

GenCore version 4.5.
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 5, 2002, 18:20:58 ; Search time 1542.23 Seconds
(without alignments)
256.727 Million cell updates/sec

Title: US-09-424-705-7

Perfect score: 24
Sequence: 1 gtagcaagcgtatgatcacc 24

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 1472140 segs, 8248589755 residues

Number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:*

- 1: gb_ba:*
- 2: gb_htg:*
- 3: gb_in:*
- 4: gb_om:*
- 5: gb_ov:*
- 6: gb_pat:*
- 7: gb_ph:*
- 8: gb_pl:*
- 9: gb_pr:*
- 10: gb_ro:*
- 11: gb_sts:*
- 12: gb_sy:*
- 13: gb_un:*
- 14: gb_vi:*
- 15: em_ba:*
- 16: em_fun:*
- 17: em_hum:*
- 18: em_in:*
- 19: em_om:*
- 20: em_ov:*
- 21: em_or:*
- 22: em_pat:*
- 23: em_ph:*
- 24: em_pl:*
- 25: em_ro:*
- 26: em_sts:*
- 27: em_sy:*
- 28: em_un:*
- 29: em_vi:*
- 30: em_htgo_hum:*
- 31: em_htgo_inv:*
- 32: em_htgo_rod:*
- 33: em_htg_hum:*
- 34: em_htg_inv:*
- 35: em_htg_rod:*
- 36: em_htg_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	24	100.0	1653	6	AX011208	AX011208 Sequence
C 2	24	100.0	1698	6	AX011206	AX011206 Sequence
C 3	22.4	93.3	418	10	D82081	D82081 Mus musculus
C 4	22.4	93.3	1568	6	E33134	E33134 Humanized a
C 5	22.4	93.3	1570	6	A77138	A77138 Sequence 6
C 6	22.4	93.3	1570	6	AR029102	AR029102 Sequence
C 7	22.4	93.3	1574	6	AX190434	AX190434 Sequence
C 8	22.4	93.3	1611	6	AX014270	AX014270 Sequence
C 9	19.4	80.8	122680	2	AL133510	AL133510 Human DNA
C 10	19.4	80.8	206773	2	AC010932	AC010932 Homo sapi
C 11	19.2	80.0	71153	9	HUAE000662	HUAE000662 Homo sapi
C 12	19.2	80.0	97630	9	HUMTCRADCV	M94081 Human Tcr-C
C 13	19.2	80.0	175053	2	AC023226	AC023226 Homo sapi
C 14	18.8	78.3	941	2	AC048636	AC048636 Giardia i
C 15	18.8	78.3	980	2	AC088680	AC088680 Giardia i
C 16	18.8	78.3	1043	2	AC048637	AC048637 Giardia i
C 17	18.8	78.3	1070	2	AC048638	AC048638 Giardia i
C 18	18.8	78.3	1094	2	AC048874	AC048874 Giardia i
C 19	18.8	78.3	101072	9	AC005752	AC005752 Homo sapi
C 20	18.8	78.3	103000	2	AC008727	AC008727 Homo sapi
C 21	18.8	78.3	169897	9	AL335674	AL335674 Human DNA
C 22	18.8	78.3	176788	2	AC021877	AC021877 Homo sapi
C 23	18.8	78.3	181576	2	AC055837	AC055837 Homo sapi
C 24	18.8	78.3	206191	2	AL359642	AL359642 Homo sapi
C 25	18.8	78.3	258654	2	AC008579	AC008579 Homo sapi
C 26	18.4	76.7	1002	2	AC053272	AC053272 Giardia i
C 27	18.4	76.7	89323	9	AP000221	AP000221 Homo sapi
C 28	18.4	76.7	100000	9	AP000084	AP000084 Homo sapi
C 29	18.4	76.7	100000	9	AP000136	AP000136 Homo sapi
C 30	18.4	76.7	171263	9	AC068243	AC068243 Homo sapi
C 31	18.4	76.7	340000	9	AP001693	AP001693 Homo sapi
C 32	18.2	75.8	740	6	A86287	A86287 Sequence 94
C 33	18.2	75.8	740	6	AR155780	AR155780 Sequence
C 34	18.2	75.8	740	6	E66305	E66305 Genome DNA
C 35	18.2	75.8	3577	10	MUST1227A	D50523 Mouse mRNA
C 36	18.2	75.8	11712	1	AE000857	AE000857 Methanoba
C 37	18.2	75.8	25660	9	AL133460	AL133460 Human DNA
C 38	18.2	75.8	94146	9	AC016604	AC016604 Homo sapi
C 39	18.2	75.8	118407	2	AC026629	AC026629 Homo sapi
C 40	18.2	75.8	122458	2	AC010391	AC010391 Homo sapi
C 41	18.2	75.8	154698	2	AC026747	AC026747 Homo sapi
C 42	18.2	75.8	157159	2	AC026279	AC026279 Homo sapi
C 43	18.2	75.8	161345	2	AC091968	AC091968 Homo sapi
C 44	18.2	75.8	162080	2	AC073872	AC073872 Homo sapi
C 45	18.2	75.8	163516	9	AC016635	AC016635 Homo sapi

ALIGNMENTS

RESULT 1
AX011208/c
LOCUS AX011208 1653 bp DNA
DEFINITION Sequence 3 from Patent WO9957150.
ACCESSION AX011208
VERSION AX011208.1 GI:9997788
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
DEUTSCHES KREBSFORSCH (DE); KIRPIYANOV SERGEY (DE); LITTLE MELVYN (DE)

FEATURES
Source
I. 1653
Location/Qualifiers
/organism="unidentified"
/db_xref="taxon:32644"

CDS

28..1647

/note="unnamed protein product"

/codon_start=1

/protein_id="CAC07543.1"

/db_xref="GI:9997789"

/translation="MKYLPPTAAAGLLLLAQAAPAAVOLOQSGAEIARPGASVKMSCKASGYFTRTMHVVKORPGGLEIMIGYINPSRGTYNOKFKKATLTIDKSSSTAYMOLSLTSEDSAVYVCARYDDHYSIDYWGQGITLVSSAKTTPKLGDDILLTPASLAVSLGRATISCKASQSVYDDSYLNWYQOIPGQPKILLIDASNLYSGIIPRFGSGSGTDTLNIHPEKVDATYHCQSTEDPWFPGGCTKLEIKADAAAGGPGSOVOLOQSGAEIARPGASVKISCKASGYFTSSYWMWVKORPGGLEIMIGIWDGDTNNGKFKKATLTADESSSTAYMOLSLTSEDSAVYVCARETTVGRYIYADYWGQTSVTVSSAKTTPKLGDDIVLTQSPALMSASPEKVTMTCSASSSVYWMYQOQSGTSPKRMIVDTSKLASGVPAHFRGSGSTSYSLTISGMEADAAATYYCCQWSSNPTFGSGTKLEINRAPTPGSPKRLISEEDLNHHHHH"

BASE COUNT

432 a 429 c 424 g 368 t

ORIGIN

Query Match

100.0%; Score 24; DB 6; Length 1653;

Best Local Similarity 100.0%; Pred. No. 0.63; Mismatches 0; Indels 0; Gaps 0;

Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gtaagcaagcgtgaatgatcatc 24

Db 417 GTAGTCAAGCGCTGAATGATCATC 394

RESULT 2

AX011206/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

AX011206 1698 bp DNA PAT 06-SEP-2000
 Sequence 1 from Patent WO957150.
 AX011206
 AX011206.1 GI:9997786

unidentified.
 unidentified
 unclassified.
 1 (bases 1 to 1698)

CDS

/note="unnamed protein product"

/protein_id="CAC07542.1"

/db_xref="GI:9997787"

/translation="MKYLPPTAAAGLLLLAQAAPAAVOLOQSGAEIARPGASVKMSCKASGYFTRTMHVVKORPGGLEIMIGYINPSRGTYNOKFKKATLTIDKSSSTAYMOLSLTSEDSAVYVCARYDDHYSIDYWGQGITLVSSAKTTPKLGDDILLTPASLAVSLGRATISCKASQSVYDDSYLNWYQOIPGQPKILLIDASNLYSGIIPRFGSGSGTDTLNIHPEKVDATYHCQSTEDPWFPGGCTKLEIKADAAAGGPGSOVOLOQSGAEIARPGASVKISCKASGYFTSSYWMWVKORPGGLEIMIGIWDGDTNNGKFKKATLTADESSSTAYMOLSLTSEDSAVYVCARETTVGRYIYADYWGQTSVTVSSAKTTPKLGDDIVLTQSPALMSASPEKVTMTCSASSSVYWMYQOQSGTSPKRMIVDTSKLASGVPAHFRGSGSTSYSLTISGMEADAAATYYCCQWSSNPTFGSGTKLEINRAPTPGSPKRLISEEDLNHHHHH"

BASE COUNT

433 a 434 c 450 g 381 t

ORIGIN

Query Match

100.0%; Score 24; DB 6; Length 1698;

Best Local Similarity 100.0%; Pred. No. 0.63; Mismatches 0; Indels 0; Gaps 0;

Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gtaagcaagcgtgaatgatcatc 24

|||||
 .|||||

Db 417 GTAGTCAAGCGCTGAATGATCATC 394

RESULT 3

D82081/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

JOURNALS

FEATURES

SOURCE

primer_bind

CDS

slg_peptide

mRNA

BASE COUNT

ORIGIN

Query Match

93.3%; Score 22.4; DB 10; Length 418;

Best Local Similarity 95.8%; Pred. No. 3.6;

Matches 23; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 gtaagcaagcgtgaatgatcatc 24

Db 384 GTAGTCAAGCGCTGAATGATCATC 361

|||||
 .|||||

|||||
 .|||||

|||||
 .|||||

|||||
 .|||||

|||||
 .|||||

|||||
 .|||||

|||||
 .|||||

|||||
 .|||||

|||||
 .|||||

|||||
 .|||||

|||||
 .|||||

|||||
 .|||||

|||||
 .|||||

AUTHORS Adia,J.R.A.A. and Singu,E.J.S.S.
TITLE Humanized antibody
JOURNAL Patent: JP 1999243955-A 14-SEP-1999;
THERATECH INC
COMMENT OS Unidentified
PN JP 1999243955-A/4
PD 14-SEP-1999
PR 22-DEC-1997 JP 1997353861
PI ADIA JOHN ROBERT,ASUMARU D SINGU, EMTUJAGE JOHN SPENCER PC
C12N15/09,A6K39/395,C07K16/46,C12N5/10//C12P21/08,C12P21/08, PC
C12R1:911,
PC C12N15/00,C12N5/00
CC Strandedness: Single;
CC Topology: Linear;
FH Key
FT source
FEATURES Location/Qualifiers
source 1..1568
/organism='Unidentified'
/db_xref='taxon:32644'
BASE COUNT 446 a 443 c 354 g 315 t 10 others
ORIGIN

Query Match 93.3%; Score 22.4; DB 6; Length 1568;
Best Local Similarity 95.8%; Pred. No.3.6;
Matches 23; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 gtagtcaagcgtgtaatgcatc 24
|||||
Db 420 GTAGTCAAGCAGTAATGATCATC 397

RESULT 5
LOCUS A77138 1570 bp DNA PAT 19-OCT-1999
DEFINITION Sequence 6 from Patent EP0620276.
ACCESSION A77138
VERSION A77138.1 GI:6088873
KEYWORDS
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 1570)
AUTHORS Adair,J.R. and Emtage,J.S.
TITLE HUMANISED ANTIBODIES
JOURNAL Patent: EP 0620276-A 6 19-OCT-1994;
CELLTECH LTD (GB)
FEATURES Location/Qualifiers
source 1..1570
/organism='unidentified'
/db_xref='taxon:32644'
BASE COUNT 447 a 445 c 354 g 324 t
ORIGIN

Query Match 93.3%; Score 22.4; DB 6; Length 1570;
Best Local Similarity 95.8%; Pred. No.3.6;
Matches 23; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 gtagtcaagcgtgtaatgcatc 24
|||||
Db 421 GTAGTCAAGCAGTAATGATCATC 398

RESULT 6
LOCUS AR029102/c 1570 bp DNA PAT 29-SEP-1999
DEFINITION Sequence 6 from patent US 5859205.
ACCESSION AR029102
VERSION AR029102.1 GI:5941075

KEYWORDS Unknown.
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1570)
AUTHORS Adair,J.Robert, Athwal,D.Singh and Emtage,J.Spencer.
TITLE Humanised antibodies
JOURNAL Patent: US 5859205-A 6 12-JAN-1999;
FEATURES Location/Qualifiers
source 1..1570
/organism='unknown'
BASE COUNT 446 a 443 c 356 g 325 t
ORIGIN

Query Match 93.3%; Score 22.4; DB 6; Length 1570;
Best Local Similarity 95.8%; Pred. No.3.6;
Matches 23; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 gtagtcaagcgtgtaatgcatc 24
|||||
Db 421 GTAGTCAAGCAGTAATGATCATC 398

RESULT 7
LOCUS AX190434/c 1574 bp DNA PAT 08-AUG-2001
DEFINITION Sequence 21 from Patent WO0147953.
ACCESSION AX190434
VERSION AX190434.1 GI:15143792
KEYWORDS
SOURCE synthetic construct.
ORGANISM synthetic construct.
REFERENCE 1 (bases 1 to 1574)
AUTHORS Mueller-Hermelink,H.K., Greiner,A., Doerken,B., Bargou,R. and
Kufner,P.
TITLE Antibodies against plasma cells
JOURNAL Patent: WO 0147953-A 21 05-JUL-2001;
Mueller-Hermelink, Hans Konrad (DE); Greiner, Axel (DE)

FEATURES Location/Qualifiers
source 1..1574
/organism='synthetic construct'
/db_xref='taxon:32630'
/note='synthetic, no natural origin'
BASE COUNT 408 a 369 c 422 g 375 t
ORIGIN

Query Match 93.3%; Score 22.4; DB 6; Length 1574;
Best Local Similarity 95.8%; Pred. No.3.6;
Matches 23; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 gtagtcaagcgtgtaatgcatc 24
|||||
Db 1143 GTAGTCAAGCAGTAATGATCATC 1120

RESULT 8
LOCUS AX014270/c 1611 bp DNA PAT 07-SEP-2000
DEFINITION Sequence 9 from Patent WO9954440.
ACCESSION AX014270
VERSION AX014270.1 GI:10040645
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 1611)
AUTHORS Bargou,R., Kufner,P., Loeffler,A. and Lutterbuese,R.
TITLE Cd19xcd3 specific polypeptides and uses thereof
JOURNAL Patent: WO 9954440-A 9 28-OCT-1999;

FEATURES
source
RIETHMUELLER GERT (DE); BARCOU RALF (DE); DOERKEN BERND (DE); KUPFER
PETER (DE); LOEFFLER ANJA (DE); LUTTERBUSE RALF (DE)
Location/Qualifiers
1. .1611
/organism="Homo sapiens"
/db_xref="taxon:9606"
11. .1606
/note="unnamed protein product"

CDS

/codon_start=1
/protein_id="CAC07568.1"
/db_xref="GI:10040646"
/translation="MGNSCTILFLVATGVSHDYKDDDKDLOLTQSPASLAVSLQ
RATISKASOSVDYDGSYLNMVQOIGQPKLLIYDASNLVSGIPRFGSGSGTGF
TLNTHPYEKYDAATVHCQSTEDPFGGKLEIKGGSGSGSGSGSOVLQDS
GAELVRPSSVKISCKASGYAFSSYMMNWKQRPQGLEMGQIIPGDDPTNYGFK
GKATLTADSSSTAYMOLSLASEDAVFCARFETTYRYAMDMVGQGTVTVS
SGGGSDIKIQSGAELRPGASVYKMSCKSGYPTRYTHMYKQRCQGLEMGYIN
PSRGTNYNOKFKDKATLTITDKSSSTAYMOLSLTSEDNAVYTCARYDDHYCIDWG
QGTTLVSVSGSGSGSGSGSGGVDIDLTQSPALMSAPGKVTMTCASSSVY
MMVYQKSGTSPKMYIDTSKVASGVPRFSGSGSGTSTLTSSMEADAAATVYCOO
MSSNPLTFGAGTKLELKHNNHH"
BASE COUNT 402 a 396 c 440 g 373 t
ORIGIN

Query Match 93.3%; Score 22.4; DB 6; Length 1611;
Best Local Similarity 95.8%; Pred. No. 3.6;
Matches 23; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OR 1 gtagtcaagctgtatgatac 24
Db 1180 GTAGTCAAGCGATGATGATCATC 1157

RESULT 9
AL133510/c
LOCUS
DEFINITION Human DNA sequence from clone RP3-481C9 on chromosome 6q25.1-26,
complete sequence.
ACCESSION AL133510
VERSION AL133510.13
KEYWORDS GI:113160149
SOURCE HTG.
ORGANISM human.
Human sapiens
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 122680)

REFERENCE 1
TITLE Direct Submission
JOURNAL Submitted (01-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
On Feb 27, 2001 this sequence version replaced g1:12717895.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.

COMMENT
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >= 30);
an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em., EMBL; SW.,
SWISSPROT; Tr., TrEMBL; Wp., WormPep; Information on the WormPep
database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping

FEATURES
source
Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr6
RP3-481C9 is from the library RPI-3 constructed by the group of
Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pCYPAC2
This sequence is the entire insert of clone RP3-481C9 The true left
end of clone RP11-193H22 is at 111676 in this sequence. The true
right end of clone RP11-52J3 is at 18402 in this sequence.
Location/Qualifiers
1. .122680
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/map="q25.1-26"
/clone="RP3-481C9"
/clone_1bp="RPI-3"
1. .235
/note="AluSc repeat: matches 59. .308 of consensus"
267. .393
/note="MER63 repeat: matches 840. .947 of consensus"
394. .700
/note="AluDb repeat: matches 1. .311 of consensus"
701. .806
/note="MER63 repeat: matches 947. .1058 of consensus"
920. .981
/note="L2 repeat: matches 1343. .1404 of consensus"
1135. .1262
/note="L2 repeat: matches 1663. .1790 of consensus"
1305. .1417
/note="L2 repeat: matches 2624. .2747 of consensus"
1419. .1874
/note="L2 repeat: matches 1855. .2317 of consensus"
1914. .2158
/note="AluJo repeat: matches 67. .309 of consensus"
2358. .2467
/note="MER91B repeat: matches 5. .119 of consensus"
2503. .2715
/note="L2 repeat: matches 2463. .2708 of consensus"
3683. .3982
/note="AluDb repeat: matches 1. .294 of consensus"
4125. .4172
/note="L2 repeat: matches 478. .520 of consensus"
4803. .5354
/note="MER41A repeat: matches 1. .554 of consensus"
5691. .6372
/note="L2 repeat: matches 1. .691 of consensus"
6904. .6931
/note="L2 repeat: matches 6128. .6154 of consensus"
6932. .7899
/note="L2 repeat: matches 1. .969 of consensus"
7910. .8177
/note="AluSc repeat: matches 28. .295 of consensus"
8178. .8702
/note="L2 repeat: matches 5582. .6120 of consensus"
8703. .9056
/note="L2 repeat: matches 1. .354 of consensus"
9057. .10256
/note="L2 repeat: matches 4358. .5586 of consensus"
10302. .10594
/note="AluSc repeat: matches 1. .301 of consensus"
10627. .10845
/note="L2 repeat: matches 34. .262 of consensus"
11043. .11153
/note="L2 repeat: matches 100. .212 of consensus"
11363. .11663
/note="AluSc repeat: matches 1. .302 of consensus"
12049. .12286
/note="L2 repeat: matches 15. .262 of consensus"
12930. .13264
/note="MER1B repeat: matches 1. .337 of consensus"
13320. .13364
/note="Alu repeat: matches 250. .294 of consensus"


```

repeat_region 14810..15003
/ote="MIR repeat: matches 20..214 of consensus"
repeat_region 15071..15381
/ote="AluSc repeat: matches 1..309 of consensus"
repeat_region 16076..16380
/ote="AluSc repeat: matches 1..303 of consensus"
repeat_region 16422..16606
/ote="LIM4 repeat: matches 3..188 of consensus"
repeat_region 17509..17584
/ote="MER5A repeat: matches 114..189 of consensus"
repeat_region 18574..18731
/ote="MER5A repeat: matches 1..161 of consensus"
repeat_region 18737..19048
/ote="AluSc repeat: matches 1..312 of consensus"
repeat_region 19141..19289
/ote="MIR repeat: matches 20..185 of consensus"
repeat_region 19832..19910
/ote="MER91C repeat: matches 38..122 of consensus"
repeat_region 20011..21487
/ote="LIPAT repeat: matches 4668..6145 of consensus"
repeat_region 21481..23974
/ote="LIPAT repeat: matches 2156..4668 of consensus"
repeat_region 24133..24427
/ote="AluSc repeat: matches 1..302 of consensus"
repeat_region 24572..24704
/ote="AluSc repeat: matches 1..134 of consensus"
repeat_region 24851..25123
/ote="AluSc repeat: matches 1..292 of consensus"
repeat_region 25857..25896
/ote="10 copies 4 mer caca 97% conserved"
repeat_region 26613..26915
/ote="AluY repeat: matches 1..302 of consensus"
repeat_region 27062..27264
/ote="LIM4 repeat: matches 5829..6031 of consensus"
repeat_region 27270..27396
/ote="FLAM_C repeat: matches 5..131 of consensus"
repeat_region 27418..27711
/ote="LIM4 repeat: matches 6018..6308 of consensus"
repeat_region 28705..28800
/ote="24 copies 4 mer ttcc 84% conserved"
repeat_region 28719..28798
/ote="2 copies 40 mer 90% conserved"
repeat_region 28823..29124
/ote="AluSc repeat: matches 1..303 of consensus"
repeat_region 29175..29474
/ote="AluSc repeat: matches 1..301 of consensus"
repeat_region 29562..29871
/ote="AluSc repeat: matches 1..310 of consensus"
repeat_region 29911..30254
/ote="LIMB repeat: matches 5829..6173 of consensus"
repeat_region 30348..30440
/ote="AluSc repeat: matches 2..97 of consensus"
repeat_region 30978..31295
/ote="AluSc repeat: matches 1..312 of consensus"
repeat_region 32699..32968
/ote="PABL_B repeat: matches 386..667 of consensus"
repeat_region 33312..33934
/ote="LIMC repeat: matches 755..1410 of consensus"
repeat_region 34198..34625
/ote="LIR1 repeat: matches 61..494 of consensus"
repeat_region 34659..35133
/ote="LIMC repeat: matches 1449..1933 of consensus"
repeat_region 35143..35232
/ote="MER2 repeat: matches 6..95 of consensus"
repeat_region 35233..35584
/ote="LIR24 repeat: matches 124..489 of consensus"
repeat_region 35717..35784
/ote="LIR23 repeat: matches 12..79 of consensus"
repeat_region 35811..36258
/ote="MER21B repeat: matches 4..486 of consensus"
repeat_region 37024..37047
/ote="12 copies 2 mer tg 95% conserved"
repeat_region 37101..37392

```

```

repeat_region 38242..38314
/ote="L2 repeat: matches 2651..2726 of consensus"
repeat_region 38541..38654
/ote="L2 repeat: matches 2097..2215 of consensus"
repeat_region 38754..38818
/ote="MER94 repeat: matches 80..134 of consensus"
repeat_region 38819..39114
/ote="AluSc repeat: matches 1..296 of consensus"
repeat_region 39115..39186
/ote="MER94 repeat: matches 5..80 of consensus"
repeat_region 39200..39283
/ote="FLAM_A repeat: matches 46..129 of consensus"
repeat_region 40048..40362
/ote="AluSc repeat: matches 1..298 of consensus"
repeat_region 41686..42301
/ote="LIPB3 repeat: matches 5506..6150 of consensus"

Query Match 80.8%; Score 19.4; DB 9; Length 122680;
Best Local Similarity 95.2%; Pred. No. 1e+02;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 tagtcaagctgtaagatca 22
Db 44609 TAATCAAGCTGTAATGATCA 44589

RESULT 10
AC010932 LOCUS
AC010932 206773 bp DNA HTG 26-MAY-2000
DEFINITION Homo sapiens chromosome 15 clone RP11-296E22 map 15, *** SEQUENCING
IN PROGRESS **, 36 unordered pieces.
ACCESSION AC010932
VERSION AC010932.3 GI:80722461
KEYWORDS HTG; HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 206773)
Birken,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 15, clone RP11-296E22
unpublished
2 (bases 1 to 206773)
Birken,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baldwin,J., Barna,N., Beckerly,R., Boguski,L., Boukhalter,B.,
Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
Cooke,P., Dearellano,K., Dewar,K., Domino,M., Donegan,L., Doyle,M.,
Ferreira,P., Fitzhugh,W., Forrest,C., Funke,R., Gage,D.,
Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howard,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Lehoczky,J., Liu,C., Locke,K., Macdonald,P., Margulis,N.,
McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrum,J.,
Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Testafave,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.

Direct Submission
Submitted (28-SEP-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On May 25, 2000 this sequence version replaced gi:6479157.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L1684
Center clone name: 296_E_22

```

* NOTE: This is a 'working draft' sequence. It currently
* consists of 36 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1375: contig of 1375 bp in length
1376 1475: gap of 100 bp
1476 3432: contig of 1957 bp in length
3433 3532: gap of 100 bp
3533 4614: contig of 1082 bp in length
4615 4714: gap of 100 bp
4715 6385: contig of 1671 bp in length
6386 6485: gap of 100 bp
6486 8036: contig of 1551 bp in length
8037 8136: gap of 100 bp
8137 9676: contig of 1540 bp in length
9677 9776: gap of 100 bp
9777 12051: contig of 2275 bp in length
12052 12151: gap of 100 bp
12152 14712: contig of 2561 bp in length
14713 14812: gap of 100 bp
14813 16605: contig of 1793 bp in length
16606 16705: gap of 100 bp
16706 18814: contig of 2109 bp in length
18815 18914: gap of 100 bp
18915 21644: contig of 2730 bp in length
21645 21744: gap of 100 bp
21745 24509: contig of 2765 bp in length
24510 24609: gap of 100 bp
24610 27483: contig of 2874 bp in length
27484 27583: gap of 100 bp
27584 28893: contig of 2310 bp in length
28894 29993: gap of 100 bp
29994 32964: contig of 2971 bp in length
32965 33064: gap of 100 bp
33065 35523: contig of 2459 bp in length
35524 35623: gap of 100 bp
35624 38490: contig of 2867 bp in length
38491 38590: gap of 100 bp
38591 43185: contig of 4595 bp in length
43186 43285: gap of 100 bp
43286 47467: contig of 4182 bp in length
47468 47567: gap of 100 bp
47568 51338: contig of 3771 bp in length
51339 51438: gap of 100 bp
51439 55273: contig of 3835 bp in length
55274 55373: gap of 100 bp
55374 58713: contig of 3340 bp in length
58714 58813: gap of 100 bp
58814 64102: contig of 5289 bp in length
64103 64202: gap of 100 bp
64203 70250: contig of 6048 bp in length
70251 70350: gap of 100 bp
70351 76887: contig of 6537 bp in length
76888 76987: gap of 100 bp
76988 84690: contig of 7703 bp in length
84691 84790: gap of 100 bp
84791 94048: contig of 9258 bp in length
94049 94148: gap of 100 bp
94149 101669: contig of 7521 bp in length
101670 101769: gap of 100 bp
101770 112318: contig of 10549 bp in length
112319 112418: gap of 100 bp
112419 121916: contig of 9498 bp in length
121917 122016: gap of 100 bp
122017 130923: contig of 8907 bp in length
130924 131023: gap of 100 bp
131024 142956: contig of 11935 bp in length
142959 143058: gap of 100 bp

FEATURES
source

143059 155724: contig of 12666 bp in length
155725 155824: gap of 100 bp
155825 167976: contig of 12152 bp in length
167977 168076: gap of 100 bp
168077 186655: contig of 18579 bp in length
186656 186755: gap of 100 bp
186756 206773: contig of 20018 bp in length.
location/Qualifiers
1. .206773
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="15"
/map="15"
/clone="RP11-296E22"
/clone_id="RPC1-11 Human Male BAC"
1. .1375
/note="assembly_fragment"
1476. .3432
/note="assembly_fragment"
3533. .4614
/note="assembly_fragment"
4715. .6385
/note="assembly_fragment"
6486. .8036
/note="assembly_fragment"
8137. .9676
/note="assembly_fragment"
9777. .12051
/note="assembly_fragment"
12152. .14712
/note="assembly_fragment"
14813. .16605
/note="assembly_fragment"
16706. .18814
/note="assembly_fragment"
18915. .21644
/note="assembly_fragment"
21745. .24509
/note="assembly_fragment"
24610. .27483
/note="assembly_fragment"
27584. .28893
/note="assembly_fragment"
29994. .32964
/note="assembly_fragment"
33065. .35523
/note="assembly_fragment"
35624. .38490
/note="assembly_fragment"
38591. .43185
/note="assembly_fragment"
43286. .47467
/note="assembly_fragment"
47568. .51338
/note="assembly_fragment"
51439. .55273
/note="assembly_fragment"
55374. .58713
/note="assembly_fragment"
clone_end:T7
vector_side:right
58814. .64102
/note="assembly_fragment"
64203. .70250
/note="assembly_fragment"
70351. .76887
/note="assembly_fragment"
clone_end:SP6
vector_side:left
76988. .84690
/note="assembly_fragment"
84791. .94048
/note="assembly_fragment"

repeat_c_region	/rpt_family="AluSx"	9599.79796
repeat_c_region	/rpt_family="MIR"	10535.10854
repeat_c_region	/rpt_family="AluY"	10835.11138
repeat_c_region	/rpt_family="AluJb"	complement(13900.14190)
repeat_c_region	/rpt_family="AluSq"	complement(20506.20810)
repeat_c_region	/rpt_family="AluY"	20928.21223
repeat_c_region	/rpt_family="AluSx"	23083.23335
repeat_c_region	/rpt_family="MIR"	25320.25618
repeat_c_region	/rpt_family="AluSq"	complement(27167.27240)
repeat_c_region	/rpt_family="MIR"	27506.27428
repeat_c_region	/rpt_family="MIR2"	complement(30158.30403)
repeat_c_region	/rpt_family="MIR"	33727.33878
repeat_c_region	/rpt_family="MIR"	complement(33920.34213)
repeat_c_region	/rpt_family="AluYb3"	complement(35275.35553)
repeat_c_region	/rpt_family="MIR2"	complement(36848.36886)
repeat_c_region	/rpt_family="L1MA3"	39321.39388
repeat_c_region	/rpt_family="MIR"	40206.40261
repeat_c_region	/rpt_family="MIR"	complement(40433.40659)
repeat_c_region	/rpt_family="MIR"	complement(40683.40848)
repeat_c_region	/rpt_family="AluSx"	complement(41322.41444)
repeat_c_region	/rpt_family="MIR"	complement(45732.45664)
repeat_c_region	/rpt_family="MIR"	46927.46992
repeat_c_region	/rpt_family="MIR2"	complement(47281.47388)
repeat_c_region	/rpt_family="MIR2"	49066.49073
repeat_c_region	/rpt_type=tandem	
repeat_c_region	/rpt_unit=rpt	49734.50032
repeat_c_region	/rpt_family="AluY"	51737.51940
repeat_c_region	/rpt_family="ME3"	complement(52892.53047)
repeat_c_region	/rpt_family="MIR"	54558.54732
repeat_c_region	/rpt_family="MIR2"	complement(64749.55052)
repeat_c_region	/rpt_family="AluSq"	55059.55302
repeat_c_region	/rpt_family="MIR2"	57276.57570
repeat_c_region	/rpt_family="AluJb"	59210.59460
repeat_c_region	/rpt_family="ME4b"	complement(61706.61769)
repeat_c_region	/rpt_family="MIR2"	63506.63580
repeat_c_region	/rpt_family="MIR"	63695.63862
repeat_c_region	/rpt_family="L1ME3"	63931.63951

```

/rpl_type=tandem
/rpl_unit=AC
repeat_region 64131..64281
/rpl_family="MIR"
repeat_region 65845..66137
/rpl_family="AluSx"
repeat_region complement(66237..66480)
/rpl_family="MIR2"
repeat_region complement(68187..68457)
/rpl_family="AluJo"
repeat_region complement(68480..68777)
/rpl_family="AluSq"
repeat_region complement(68778..68837)
/rpl_family="FAM"
repeat_region 69326..69628
/rpl_family="AluSp"
repeat_region complement(69688..69766)
/rpl_family="MER58A"
repeat_region 69845..69897
/rpl_family="MER5A"
repeat_region complement(69907..70048)
/rpl_family="MER5A"
repeat_region complement(70060..70225)
/rpl_family="MER5B"
BASE COUNT 20125 a 15457 c 16410 g 19161 t
ORIGIN

```

```

Query Match 80.0%; Score 19.2; DB 9; Length 71153;
Best Local Similarity 87.5%; Pred. NO. 1.3e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

Oy 1 gtatgcaggctgaatgcatc 24
|||||
Db 33834 GTAGTCAGGCTGCAATAATATC 33811

```

```

RESULT 12
HUMTCRADCV/c HUMTCRADCV 97630 bp DNA PRI 21-NOV-1997
LOCUS Human Tcr-C-delta gene, exons 1-4; Tcr-V-delta gene, exons 1-2;
DEFINITION T-cell receptor alpha (Tcr-alpha) gene, J1-J61 segments; and
Tcr-C-alpha gene, exons 1-4.

```

```

ACCESSION M94081.1 GI:2627263
VERSION M94081
KEYWORDS T-cell receptor C-alpha; T-cell receptor C-delta; T-cell receptor
V-delta; T-cell receptor alpha.

```

```

RCE human.
RGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 97630)
AUTHORS Koop,B.F., Rowen,L., Seto,D., Kuo,C.-I., Hood,L. and Wang,K.
TITLE Nucleotide sequence of the 3' terminal end of the T-cell receptor
alpha/delta locus
JOURNAL Unpublished (1992)
COMMENT On Nov 20, 1997 this sequence version replaced gi:339242.
FEATURES
source location/Qualifiers
1..97630
/organism="Homo sapiens"
/db_xref="taxon:9606"
/germline
3168..3464
/gene="Tcr-C-delta"
/number=1
3168..6809
/gene="Tcr-C-delta"
4028..4098
/gene="Tcr-C-delta"
/number=2
4458..4572
/gene="Tcr-C-delta"
/number=3

```

```

exon
gene
exon
exon

```

```

exon 5787..6809
/gene="Tcr-C-delta"
/gene="Tcr-C-delta"
9295..9868
/gene="Tcr-V-delta"
9295..9598
/gene="Tcr-V-delta"
/number=1
9832..9868
/gene="Tcr-V-delta"
/number=2
<15555..>85296
/gene="Tcr-alpha"
<15555..>15627
/gene="Tcr-alpha"
/standard_name="possible J 61 gene segment"
/note="putative"
/codon_start=3
/protein_id="AAB86739.1"
/db_xref="GI:553731"
/translation="SCGYRVNRKLTFGANTRGIMKLS"
16556..16610
/gene="Tcr-alpha"
/standard_name="possible J 60 gene segment"
/note="putative"
<16801..>16850
/gene="Tcr-alpha"
/standard_name="possible J 59 gene segment"
/note="putative"
/codon_start=2
/protein_id="AAB86740.1"
/db_xref="GI:553732"
/translation="EGRKQFIYINWDASE"
17953..18016
/gene="Tcr-alpha"
/standard_name="possible J 58 gene segment"
/note="putative"
<19121..>19181
/gene="Tcr-alpha"
/standard_name="possible J 57 gene segment"
/note="putative"
/codon_start=1
/protein_id="AAB86741.1"
/db_xref="GI:553733"
/translation="TQCGSEKLVGKGTKLTVNP"
<19765..>19829
/gene="Tcr-alpha"
/standard_name="possible J 56 gene segment"
/note="putative"
/codon_start=2
/protein_id="AAB86742.1"
/db_xref="GI:553734"
/translation="CYTGANSKLTFGKGTILTVNP"
<22523..>22593
/gene="Tcr-alpha"
/standard_name="possible J 55 gene segment"
/note="putative"
/codon_start=1
/protein_id="AAB86743.1"
/db_xref="GI:553735"
/translation="VLCNNGSPGAGIWPNOADYCP"
<23251..>23316
/gene="Tcr-alpha"
/standard_name="possible J 54 gene segment"
/note="putative"
/codon_start=3
/protein_id="AAB86744.1"
/db_xref="GI:553736"
/translation="NSGGSNYKLTFGKGTILTVNP"
<26473..>26542
/gene="Tcr-alpha"
/standard_name="possible J 53 gene segment"
/note="putative"

```

```

CDS
CDS
CDS

```

exon

/codon_start=1
/protein_id="AAB86745.1"
/db_xref="GI:553737"
/translation="ANAGTSTYKLFPGQGITLVHP"
27429..27491
/partial
/gene="Tcr-alpha"
/standard_name="possible J 52 gene segment"
/note="putative"
/db_xref="GI:27905"
/gene="Tcr-alpha"
/standard_name="possible J 51 gene segment"
/note="putative"
/codon_start=3
/protein_id="AAB86746.1"
/db_xref="GI:339250"
/translation="KSLYFWSSCCACIVQADST"
28839..28895
/gene="Tcr-alpha"
/standard_name="possible J 50 gene segment"
/note="putative"
/codon_start=2
/protein_id="AAB86747.1"
/db_xref="GI:339251"
/translation="ENLLKQGDWARDKLSHS"
29734..29789
/gene="Tcr-alpha"
/standard_name="possible J 49 gene segment"
/note="putative"
/codon_start=2
/protein_id="AAB86748.1"
/db_xref="GI:553738"
/translation="NTGNQFYFGTSLTVLP"
30737..30799
/gene="Tcr-alpha"
/standard_name="possible J 48 gene segment"
/note="putative"
/codon_start=3
/protein_id="AAB86749.1"
/db_xref="GI:553739"
/translation="SNKGNKLTFGTGLTLP"
33096..33152
/gene="Tcr-alpha"
/standard_name="possible J 47 gene segment"
/note="putative"
/codon_start=2
/protein_id="AAB86750.1"
/db_xref="GI:339254"
/translation="GIMKQKGLMRHNSQVL"
33647..33709
/gene="Tcr-alpha"
/standard_name="possible J 46 gene segment"
/note="putative"
/codon_start=3
/protein_id="AAB86751.1"
/db_xref="GI:553740"
/translation="KKSSGDKLTFGTGTRLAVRP"
34168..34234
/gene="Tcr-alpha"
/standard_name="possible J 45 gene segment"
/note="putative"
/codon_start=1
/protein_id="AAB86752.1"
/db_xref="GI:553741"
/translation="VYSGGADGLTFGKTHLITP"
35064..35126
/gene="Tcr-alpha"
/standard_name="possible J 44 gene segment"
/note="putative"
/codon_start=3
/protein_id="AAB86753.1"
/db_xref="GI:553742"
/translation="NTGTASKLTFGTGRLQVTL"

CDS

<36139..36207
/gene="Tcr-alpha"
/standard_name="possible J 43 gene segment"
/note="putative"
/codon_start=3
/protein_id="AAB86754.1"
/db_xref="GI:553743"
/translation="MYCDNNMDRFAGTRLTVPK"
37130..37195
/gene="Tcr-alpha"
/standard_name="possible J 42 gene segment"
/note="putative"
/codon_start=3
/protein_id="AAB86755.1"
/db_xref="GI:553744"
/translation="NYGSGGNLTFGKTKLSVKP"
37899..37961
/gene="Tcr-alpha"
/standard_name="possible J 41 gene segment"
/note="putative"

Query Match 80.0%; Score 19.2; DB 9; Length 97630;
Best Local Similarity 87.5%; Pred. No. 1.3e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 gtagcaagctgtaatgcatc 24
|||||
Db 67943 GTACTCAAGCTGGAATATATATC 67920

RESULT 13
AC023226/C
LOCUS
DEFINITION Homo sapiens chromosome 14 clone RP11-576P2 map 14, WORKING DRAFT
SEQUENCE, 21 unordered pieces.
ACCESSION AC023226.3 GI:7145013
VERSION
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS

1 (bases 1 to 175053)
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
Birren, B., Linton, L., Nusbaum, C., and Lander, E.
Homo sapiens chromosome 14, clone RP11-576P2
2 (bases 1 to 175053)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
Anderson, S., Baldwin, J., Barna, N., Beckerly, R., Beda, F.,
Boguslavsky, L., Boukhgalter, B., Brown, A., Burkett, G., Castle, A.,
Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P.,
DeArrelano, K., Dewar, K., Domino, M., Doyle, M., Fenesfor, J.,
Ferreira, P., Fitzhugh, M., Forrest, C., Gage, D., Galagan, J.,
Gardyna, S., Grant, G., Hagos, B., Heald, A., Horton, L.,
Howard, J. C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J.,
Lander, T., Lehoczy, J., Levine, R., Lieu, C., Liu, G., Locke, K.,
Macdonald, P., Marquis, N., McEwan, P., McGurk, A., McKernan, K.,
Mcneeters, R., Meldrum, J., Menes, L., Morrow, J., Naylor, J.,
Norman, C. H., O'Connor, T., O'Donnell, P., Oliver, T. M., Peterson, K.,
Pierre, N., Pisan, C., Pollara, V., Raymond, C., Riley, R., Rothman, D.,
Roy, A., Santos, R., Severy, P., Spencer, B., Stange-Thomann, N.,
Stojanovic, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,
Tirrell, A., Vassiliev, H., Viel, R., Vo, A., Wu, X., Wyman, D., Ye, W. J.,
Zimmer, A. and Zody, M.

TITLE
JOURNAL
COMMENT

Submitted (09-FEB-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 3, 2000 this sequence version replaced gi:7139832.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

Center: Whitehead Institute/ MIT Center for Genome Research

```

Center code: MIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
-----
Project Information
Center project name: 576_P-2
Center clone name: 576_P-2
-----
Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 161058 bases at least Q40
Consensus quality: 167879 bases at least Q30
Consensus quality: 170671 bases at least Q20
Insert size: 182000; agarose-fp
Insert size: 173053; sum-of-contigs
Quality coverage: 3.8 in Q20 bases; agarose-fp
Quality coverage: 4.0 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 21 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
1 1065: contig of 1065 bp in length
* 1066 1165: gap of 100 bp
* 1166 2413: contig of 1248 bp in length
* 2414 2513: gap of 100 bp
* 2514 4839: contig of 2326 bp in length
* 4840 4939: gap of 100 bp
* 4940 6965: contig of 2026 bp in length
* 6966 7065: gap of 100 bp
* 7066 8841: contig of 1776 bp in length
* 8842 8941: gap of 100 bp
* 8942 11475: contig of 2534 bp in length
* 11476 11575: gap of 100 bp
* 11576 14560: contig of 2985 bp in length
* 14561 14660: gap of 100 bp
* 14661 19338: contig of 4678 bp in length
* 19339 19438: gap of 100 bp
* 19439 22444: contig of 3006 bp in length
* 22445 22544: gap of 100 bp
* 22545 27238: contig of 4694 bp in length
* 27239 27338: gap of 100 bp
* 27339 33032: contig of 5694 bp in length
* 33033 33132: gap of 100 bp
* 33133 38008: contig of 4876 bp in length
* 38009 38108: gap of 100 bp
* 38109 43595: contig of 5487 bp in length
* 43596 43695: gap of 100 bp
* 43696 51575: contig of 7880 bp in length
* 51576 51675: gap of 100 bp
* 51676 59667: contig of 7992 bp in length
* 59668 59767: gap of 100 bp
* 59768 69201: contig of 9434 bp in length
* 69202 69301: gap of 100 bp
* 69302 79384: contig of 10083 bp in length
* 79385 79484: gap of 100 bp
* 79485 91749: contig of 12265 bp in length
* 91750 91849: gap of 100 bp
* 91850 112964: contig of 21115 bp in length
* 112965 113064: gap of 100 bp
* 113065 134547: contig of 21483 bp in length
* 134548 134647: gap of 100 bp
* 134648 175053: contig of 40406 bp in length.
Location/Qualifiers
1. 175053
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="14"
/map="14"

```

```

/clone="Rp11-576P2"
/clone.lib="RPC1-11 Human Male BAC"
1. 1065
misc_feature
/note="assembly_fragment"
1166. 2413
/note="assembly_fragment"
2514. 4839
/note="assembly_fragment"
4940. 6965
/note="assembly_fragment"
7066. 8841
/note="assembly_fragment"
8942. 11475
/note="assembly_fragment"
11576. 14560
/note="assembly_fragment"
14661. 19338
/note="assembly_fragment"
19439. 22444
/note="assembly_fragment"
22545. 27238
/note="assembly_fragment"
clone_end:77
vector_side:left"
27339. 33032
/note="assembly_fragment"
33133. 38008
/note="assembly_fragment"
38109. 43595
/note="assembly_fragment"
clone_end:566
vector_side:right"
43696. 51575
/note="assembly_fragment"
51676. 59667
/note="assembly_fragment"
59768. 69201
/note="assembly_fragment"
69302. 79384
/note="assembly_fragment"
79485. 91749
/note="assembly_fragment"
91850. 112964
/note="assembly_fragment"
113065. 134547
/note="assembly_fragment"
134648. 175053
/note="assembly_fragment"
BASE COUNT 49542 a 38101 c 38584 g 46814 t 2012 others
ORIGIN
Query Match 80.0%; Score 19.2; DB 2; Length 175053;
Best Local Similarity 87.5%; Pred. No. 1.3e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Oy 1 gtagcgaagctgtaataatc 24
db 88787 GTAGTCAAGCTGGAATATATC 88764
RESULT 14
AC048636/c 941 bp DNA 14-APR-2000
DEFINITION Giardia intestinalis clone KJ1200 strain WB-C6, LOW-PASS SEQUENCE
ACCESSION AC048636
VERSION AC048636.1 GI:7550065
KEYWORDS HTG; HTGS; PHASEO
SOURCE Giardia intestinalis
ORGANISM Eukaryota; Diplomonadida; Hexamitidae; Giardia;
REFERENCE 1 (bases 1 to 941)

```

AUTHORS
Morrison,H.G., McArthur,A.G., Nixon,J., Eakin,N.O., Kim,U.,
Crocker,M.C., Hinkle,G., Holder,M.E. and Sogin,M.L.
Giardia: a model for ancient eukaryotic genome analysis
JOURNAL
Unpublished
REFERENCE
2 (bases 1 to 941)
AUTHORS
Kim,U., Morrison,H.G., McArthur,A.G., Nixon,J., Eakin,N.O.,
Crocker,M.C., Hinkle,G., Holder,M.E. and Sogin,M.L.
TITLE
Direct Submission
JOURNAL
Submitted (14-APR-2000) Josephine Bay Paul Center for Comparative
Molecular Biology and Evolution, Marine Biological Laboratory, 7
MBL Street, Woods Hole, MA 02543-1015, USA
COMMENT
* NOTE: This record contains 1 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
1
Location/Qualifiers
1. 941
/organism="Giardia intestinalis"
/strain="WB-C6"
/db_xref="taxon:5741"
/clone="KJ1200"
BASE COUNT 251 a 226 c 221 g 243 t
ORIGIN

Query Match 78.3% Score 18.8; DB 2: Length 941;
Best Local Similarity 90.9%; Pred. No. 1.9e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gtagtcaagctgtaatgatca 22
|||||
Db 259 GTAGTCAAGCTATTAAGATCA 238

RESULT 15
AC088680
LOCUS
DEFINITION
Giardia intestinalis clone KJ6211 strain WB-C6, LOW-PASS SEQUENCE
SAMPLING.
AC088680 980 bp DNA HTG 06-FEB-2001
AC088680.1 GI:12668594
HTG; HTGS-PHASE0.
ORGANISM
Giardia intestinalis.
Eukaryota; Diplomonadida; Hexamitidae; Giardinae; Giardia.
1 (bases 1 to 980)
Morrison,H.G., McArthur,A.G., Nixon,J., Eakin,N.O., Kim,U.,
Crocker,M.C., Hinkle,G., Holder,M.E. and Sogin,M.L.
Giardia: a model for ancient eukaryotic genome analysis
JOURNAL
Unpublished
REFERENCE
2 (bases 1 to 980)
AUTHORS
Nixon,J., Morrison,H.G., McArthur,A.G., Eakin,N.O., Kim,U.,
Crocker,M.C., Hinkle,G., Holder,M.E. and Sogin,M.L.
TITLE
Direct Submission
JOURNAL
Submitted (06-FEB-2001) Josephine Bay Paul Center for Comparative
Molecular Biology and Evolution, Marine Biological Laboratory, 7
MBL Street, Woods Hole, MA 02543-1015, USA
COMMENT
* NOTE: This record contains 1 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone

* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
1
Location/Qualifiers
1. 980
/organism="Giardia intestinalis"
/strain="WB-C6"
/db_xref="taxon:5741"
/clone="KJ6211"
BASE COUNT 255 a 239 c 227 g 259 t
ORIGIN

Query Match 78.3% Score 18.8; DB 2: Length 980;
Best Local Similarity 90.9%; Pred. No. 1.9e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gtagtcaagctgtaatgatca 22
|||||
Db 943 GTAGTCAAGCTATTAAGATCA 964

Search completed: March 5, 2002, 19:15:06
Job time: 3248 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Comugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 5, 2002, 18:19:58 ; Search time 1530.11 Seconds

(without alignments)
168.549 Million cell updates/sec

US-09-424-705-7

Title: 24
Perfect score: 1 gtagcaagctgtaatgatac 24
Sequence:

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 11351937 segs, 5372889281 residues

number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: em_estfun:*
2: em_esthum:*
3: em_estin:*
4: em_estom:*
5: em_estpl:*
6: em_estdb:*
7: em_estro:*
8: em_estov:*
9: em_hic:*
10: qb_est1:*
11: qb_est2:*
12: qb_hic:*
13: qb_gss:*
14: em_gss_fun:*
15: em_gss_hum:*
16: em_gss_huv:*
17: em_gss_pin:*
18: em_gss_pro:*
19: em_gss_tod:*
20: em_gss_vit:*
21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	18.8	78.3	490	13	AO121346 HS.3083.B
C 2	18.8	78.3	613	13	AO457569 HS.5097.B
C 3	18.8	78.3	614	13	CNS02BEA
C 4	18.4	76.7	702	10	AM442262 EST311658
C 5	18.2	75.8	436	11	B1319816
C 6	18.2	75.8	471	11	B1319816
C 7	18.2	75.8	471	11	B1319816
C 8	18.2	75.8	471	11	B1319816
C 9	18.2	75.8	511	13	AO831679 HS.2089.A
C 10	18.2	75.8	511	13	AO713927 HS.2122.B
C 11	18.2	75.8	513	11	B1338823
C 12	18.2	75.8	550	11	B1333300 UI-M-BH3-BE948941

13	18.2	75.8	603	11	B1134657	B1134657	UI-M-BH3-
14	18.2	75.8	610	11	BE949315	BE949315	UI-M-BH3-
C 15	18.2	75.8	653	11	BE915494	602815718	BE915494
16	18.2	75.8	811	11	BE873814	601483965	BE873814
17	18.2	75.8	882	10	BE747423	601575514	BE747423
18	18.2	75.8	919	11	BE346836	602021638	BE346836
19	18.2	75.8	985	10	AL518879	602021638	AL518879
20	18.2	75.8	1010	11	BE479767	602326973	BE479767
C 21	18.2	75.8	1111	11	BE869211	601445017	BE869211
C 22	18.2	75.8	1176	10	BE282889	601100528	BE282889
C 23	18.2	75.8	1885	11	BE850184	102402881	BE850184
C 24	17.8	74.2	278	10	A1632464	wb09406.x	A1632464
C 25	17.8	74.2	278	10	AM954601	hg67d06.x	AM954601
C 26	17.8	74.2	562	10	AM829023	ra70h07.y	AM829023
C 27	17.8	74.2	1176	11	BC032792	602298454	BC032792
C 28	17.6	73.3	329	10	BE244322	TCBAP2E03	BE244322
C 29	17.6	73.3	359	11	BE378808	IL2-UM008	BE378808
30	17.6	73.3	388	10	AM803587	IL2-UM008	AM803587
31	17.6	73.3	513	10	AM455383	UI-M-BH2.	AM455383
32	17.6	73.3	549	13	AO988825	13ELE01NE	AO988825
33	17.6	73.3	582	13	AO021176	CIT-HSP-2	AO021176
34	17.6	73.3	598	10	AA871135	vg31e06.r	AA871135
35	17.6	73.3	663	13	A2872517	2M0246422	A2872517
36	17.6	73.3	688	13	A2898070	RPCT-24-1	A2898070
C 37	17.6	73.3	761	13	AO935036	CPG2380A	AO935036
C 38	17.6	73.3	766	11	B1145485	602910161	B1145485
C 39	17.6	73.3	796	13	CNS03DRO	Tetradon	AL339471
C 40	17.6	73.3	846	11	B1144614	602909496	B1144614
C 41	17.6	73.3	849	11	BE272799	GA_Eb001	BE272799
C 42	17.6	73.3	852	11	BE272801	GA_Eb001	BE272801
C 43	17.6	73.3	853	11	BE272797	GA_Eb001	BE272797
C 44	17.6	73.3	865	11	B1247100	602960665	B1247100
C 45	17.4	72.5	221	11	BE956145	QV2-NN004	BE956145

ALIGNMENTS

RESULT 1
LOCUS AO121346 490 bp DNA GSS 22-SEP-1998
DEFINITION HS_3083_B1.D10.MF CIT Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate=3083 Col=19 Row=H, DNA sequence.

ACCESSION AO121346
VERSION AO121346.1 GI:3498512
KEYWORDS GSS.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 490)
Mahalas, G.G., Wallace, J.C., Smith, K., Swartzell, S., Holzman, T., Keller, A., Shaker, R., Furlong, V., Young, V., Zhao, S., Adams, M.D. and Hood, L.

TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome

JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
COMMENT 99380589

CONTACT: Mahalas GG, Wallace JC, Hood L

High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887

Email: jwallace@u.washington.edu
Sequence Tagged Connector
Plate: 3083 Row: H Column: 19

Class: BAC ends
High quality sequence stop: 490.

Location/Qualifiers
1..490
/organism="Homo sapiens"
/db_xref="taxon:9606"

/clone="plate=3083 Col=19 Row=H"
/sex="male"
/note="Organ: sperm; Vector: pBelOBAC11; BAC clones in E-Coli DH10B"

BASE COUNT 117 a 99 c 95 g 174 t 5 others

ORIGIN

Query Match 78.3%; Score 18.8; DB 13; Length 490;
Best Local Similarity 90.9%; Pred. No. 2.9e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 gtagtcaagctgtaatgac 22
|||||
Db 364 GTAGTCAAGCTGCAATGACGA 343

RESULT 2
US 67569/c
DEFINITION AQ457569 613 bp DNA GSS 23-APR-1999
HS_5097.B1.D07.T7A RPCI-11 Human Male BAC Library Homo sapiens
genomic clone Plate=673 Col=13 Row=H, DNA sequence.
ACCESSION AQ457569
VERSION AQ457569.1 GI:4636209
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 613)
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.
Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
99380589
CONTACT: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones are derived from the human BAC library RPCI-11. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.bufileo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.bufileo.edu/ordering_bac.htm)
or from Research Genetics (info@resgen.com). BAC end Web Server:
http://www.htsc.washington.edu
Plate: 673 Row: H Column: 13
Seq primer: T7
Class: BAC ends
High quality sequence stop: 613.

FEATURES

source

1. 613
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="Plate=673 Col=13 Row=H"
/clone="lib="RPCI-11 Human Male BAC Library"
/sex="male"
/note="Vector: pBAC3.6; Site_1: EcoRI; Site_2: EcoRI;
Male blood DNA was isolated from one randomly chosen donor
and partially digested with a combination of EcoRI and
EcoRI Methylase. Size selected DNA was cloned into the
pBAC3.6 vector at EcoRI sites"
BASE COUNT 141 a 142 c 107 g 202 t 21 others

Query Match 78.3%; Score 18.8; DB 13; Length 613;
Best Local Similarity 90.9%; Pred. No. 2.9e+02;

Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 1 gtagtcaagctgtaatgac 22
|||||
Db 424 GTAGTCAAGCTGCAATGACGA 403

RESULT 3
CNS02EEA 614 bp DNA GSS 13-MAY-2000
LOCUS Tetradon nigroviridis genome survey sequence PUC-ori end of clone
DEFINITION 262017 of library G from Tetradon nigroviridis, genomic survey
sequence.
ACCESSION AL193627.1 GI:7831733
VERSION AL193627.1 GI:7831733
KEYWORDS GSS; genome survey sequence.
SOURCE Tetradon nigroviridis.
ORGANISM Tetradon nigroviridis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percormorpha; Tetraodontiformes;
Tetraodontidae; Tetraodon.

REFERENCE 1 (bases 1 to 614)
AUTHORS Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
Weissenbach,J.
Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetraodon nigroviridis

TITLE Unpublished
JOURNAL 2 (bases 1 to 614)
REFERENCE Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
AUTHORS Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
Saurin,W. and Weissenbach,J.
Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
Unpublished
3 (bases 1 to 614)
Genoscope.

JOURNAL Direct Submission
TITLE Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases
REFERENCE This sequence is a single read and was generated as part of a large
AUTHORS scale clone-end sequencing project of the Tetraodon nigroviridis
COMMENT genome. For more information, please take a look at
http://www.genoscope.cns.fr/tetraodon.

FEATURES

source

1. 614
Location/Qualifiers
/organism="Tetraodon nigroviridis"
/db_xref="taxon:99883"
/clone="262017"
/clone_lib="G"
/note="Genoscope sequence ID : C0AG262AH09SP1-end ;
PUC-ori"
BASE COUNT 154 a 164 c 105 g 160 t 31 others

ORIGIN

Query Match 78.3%; Score 18.8; DB 13; Length 614;
Best Local Similarity 90.9%; Pred. No. 2.9e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 3 agtcaagctgtaatgac 24
|||||
Db 217 AGTCAAGCTGTAATGACATC 238

RESULT 4
AM442262 702 bp mRNA EST 18-MAY-2001
LOCUS EST311658 tomato fruit red ripe, TMU Lycopersicon esculentum cDNA
DEFINITION EST311658 tomato fruit red ripe, TMU Lycopersicon esculentum cDNA
ACCESSION AM442262
VERSION AM442262.1 GI:6977513
KEYWORDS EST.

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 5, 2002, 18:22:38 ; Search time 67.11 Seconds
(Without alignments)
80.993 Million cell updates/sec

Title: US-09-424-705-7
Perfect score: 24
Sequence: 1 gtagcaagctgtatgatcatc 24

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Issued_Patents_NA: *
1: /cgn2_6/prodata/1/ina/5A-COMB.seq: *
2: /cgn2_6/prodata/1/ina/5B-COMB.seq: *
3: /cgn2_6/prodata/1/ina/6A-COMB.seq: *
4: /cgn2_6/prodata/1/ina/6B-COMB.seq: *
5: /cgn2_6/prodata/1/ina/PCTUS-COMB.seq: *
6: /cgn2_6/prodata/1/ina/backfiles1.seq: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	22.4	93.3	1570	2	US-08-303-569B-6
C 2	22.4	93.3	1570	2	US-08-116-247-6
C 3	18.2	75.8	740	4	US-08-998-416-946
C 4	16.6	69.2	2297	1	US-08-056-200-110
C 5	16.6	69.2	2297	2	US-08-800-644-110
C 6	16.6	69.2	2297	4	US-08-991-789A-183
C 7	16.6	69.2	2297	4	US-08-998-416-569
C 8	16.6	69.2	2297	4	US-08-998-416-569
C 9	16.6	69.2	2297	4	US-08-998-416-569
C 10	16.6	69.2	2297	4	US-08-998-416-569
C 11	15.6	65.0	1610	4	US-08-387-707-13
C 12	15.6	65.0	1610	4	US-08-889-108-7
C 13	15.6	65.0	1610	5	PCT-US94-10358-7
C 14	15.6	65.0	2038	2	US-08-736-770-2
C 15	15.6	65.0	2038	2	US-08-188-275A-1
C 16	15.6	65.0	2038	2	US-08-188-275A-1
C 17	15.2	63.3	176373	4	US-09-128-155-17
C 18	15.2	63.3	176373	4	US-09-385-982-89
C 19	15.2	63.3	176373	4	US-09-385-982-89
C 20	15.2	63.3	3500	4	US-09-197-636-7
C 21	15.2	63.3	4164	1	US-08-188-582-3
C 22	15.2	63.3	4803	4	US-08-646-715-3
C 23	15.2	63.3	4803	4	US-09-197-636-1
C 24	15.2	63.3	246240	2	US-08-724-394A-20
C 25	15.2	63.3	246240	2	US-08-724-394A-21
C 26	15.2	63.3	246240	2	US-08-724-394A-22
C 27	15.2	63.3	4411529	4	US-09-103-840A-1

C 28	15	62.5	357	1	US-08-318-970B-31	Sequence 31, Appl
C 29	15	62.5	399	1	US-08-318-970B-40	Sequence 40, Appl
C 30	15	62.5	402	1	US-08-318-970B-41	Sequence 41, Appl
C 31	15	62.5	1210	2	US-08-483-695-4	Sequence 4, Appl
C 32	15	62.5	1210	2	US-07-965-285-4	Sequence 4, Appl
C 33	15	62.5	1210	2	US-08-487-231-4	Sequence 4, Appl
C 34	15	62.5	1210	4	US-09-201-912-4	Sequence 1, Appl
C 35	15	62.5	1260	3	US-08-789-982-1	Sequence 1, Appl
C 36	15	62.5	1761	3	US-08-861-747-1	Sequence 1, Appl
C 37	15	62.5	1889	3	US-08-861-747-3	Sequence 3, Appl
C 38	15	62.5	1894	1	US-08-476-008-66	Sequence 66, Appl
C 39	15	62.5	1894	1	US-08-306-063-66	Sequence 66, Appl
C 40	15	62.5	1894	1	US-08-833-485-66	Sequence 66, Appl
C 41	15	62.5	1894	4	US-09-137-440-66	Sequence 66, Appl
C 42	15	62.5	2185	2	US-08-467-948A-3	Sequence 3, Appl
C 43	15	62.5	2185	3	US-08-467-947A-3	Sequence 3, Appl
C 44	15	62.5	2355	4	US-08-913-159-12	Sequence 12, Appl
C 45	15	62.5	2463	1	US-08-370-542-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1
US-08-303-569B-6/c
Sequence 6, Application US/08303569B
Patent No. 5859205
GENERAL INFORMATION:
APPLICANT: Adair, John R.
APPLICANT: Athwal, Diljeet S.
APPLICANT: Emtage, John S.
TITLE OF INVENTION: Humanised Antibodies
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESS:
ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 5859205RIS
STREET: One Liberty Place - 46th Floor
CITY: Philadelphia
STATE: PA
COUNTRY: USA
ZIP: 19103
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/303,569B
FILING DATE: 07-SEP-1994
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Trujillo, Doreen Yalco
REGISTRATION NUMBER: 35,719
REFERENCE/DOCKET NUMBER: CARP-0032
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 568-3100
TELEFAX: (215) 568-3439
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 1570 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 41..1444
FEATURE:
NAME/KEY: mat.peptide
LOCATION: 98..1444
US-08-303-569B-6
Query Match 93.3%; Score 22.4; DB 2; Length 1570;

Best Local Similarity 95.8%; Pred. No. 0.027;
Matches 23; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 gtagtcaagcgtgaatgcatc 24
|||||
Db 421 GTACTCAAGCGAGTAATGATCATC 398

RESULT 2

US-08-116-247-6/C
Sequence 6, Application US/08116247
Patent No. 5929212
GENERAL INFORMATION:
APPLICANT: Jolliffe, Linda K.
APPLICANT: Zivin, Robert A.
APPLICANT: Athwal, Diljeet S.
TITLE OF INVENTION: CD3 Specific Recombinant Antibody
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: Woodcock Masburn Kurtz Mackiewicz & No. 5929212iris
STREET: One Liberty Place - 46th Floor
CITY: Philadelphia
STATE: PA
COUNTRY: USA
ZIP: 19103

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/116,247
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/743,377
FILING DATE: 10-OCT-1991
ATTORNEY/AGENT INFORMATION:
NAME: Painulin, Francis A.
REGISTRATION NUMBER: 19,386
REFERENCE/DOCKET NUMBER: CARP-0011
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 568-3100
TELEFAX: (215) 568-3439
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 1570 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 41..1444
US-08-116-247-6

Query Match 93.3%; Score 22.4; DB 2; Length 1570;
Best Local Similarity 95.8%; Pred. No. 0.027;
Matches 23; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 gtagtcaagcgtgaatgcatc 24
|||||
Db 421 GTACTCAAGCGAGTAATGATCATC 398

RESULT 3

US-08-998-416-946/C
Sequence 946, Application US/08998416
Patent No. 6239264
GENERAL INFORMATION:
APPLICANT: Philippsen, Peter
APPLICANT: Pohlmann, Rainer

APPLICANT: Steiner, Sabine
APPLICANT: Mohr, Christine
APPLICANT: Wendland, Jürgen
APPLICANT: Knechtle, Philipp
APPLICANT: Reibschung, Corinne
TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSYPPII
NUMBER OF SEQUENCES: 1152
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 6239264artis Corporation
STREET: 3054 Cornwallis Road
CITY: Research Triangle Park
STATE: No. 6239264th Carolina
COUNTRY: USA
ZIP: 27709

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/998,416
FILING DATE: 24-DEC-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: CH 0016/97
FILING DATE: 31-DEC-1996
ATTORNEY/AGENT INFORMATION:
NAME: Meigs, J. Timothy
REGISTRATION NUMBER: 38,241
REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8587
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 946:
SEQUENCE CHARACTERISTICS:
LENGTH: 740 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: PAG1585UP
US-08-998-416-946

Query Match 75.8%; Score 18.2; DB 4; Length 740;
Best Local Similarity 87.0%; Pred. No. 2.8;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 gtagtcaagcgtgaatgcatc 23
|||||
Db 172 GTAGTCAGCCTGTATGATTAAT 150

RESULT 4

US-08-056-200-110/C
Sequence 110, Application US/08056200
Patent No. 5616500
GENERAL INFORMATION:
APPLICANT: Steinert, Peter M.
APPLICANT: Lee, Seung-Chul
APPLICANT: Kim, In-Gyu
APPLICANT: Chung, Soo-Il
APPLICANT: Park, Sang-Chul
TITLE OF INVENTION: Trichothyalin and Transglutaminase-3 and
NUMBER OF SEQUENCES: 117
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knobbe, Martens, Olson & Bear
STREET: 620 Newport Center Drive, Sixteenth Floor
CITY: Newport Beach
STATE: CA

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 5, 2002, 19:15:08 ; Search time 1556.36 Seconds
(without alignments)
254.396 Million cell updates/sec

Title: US-09-424-705-7

Perfect score: 24
1 gtagcaagcgtcaatgcatc 24

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

number of hits satisfying chosen parameters: 309068

Minimum DB seq length: 0
Maximum DB seq length: 24

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenBank: *
1: gb_ba: *
2: gb_htg: *
3: gb_in: *
4: gb_cm: *
5: gb_ov: *
6: gb_pat: *
7: gb_ph: *
8: gb_pl: *
9: gb_pr: *
10: gb_ro: *
11: gb_sts: *
12: gb_sy: *
13: gb_un: *
14: gb_vi: *
15: em_ba: *
16: em_fun: *
17: em_hum: *
18: em_in: *
19: em_om: *
20: em_or: *
21: em_ov: *
22: em_pat: *
23: em_ph: *
24: em_pl: *
25: em_ro: *
26: em_sts: *
27: em_sy: *
28: em_un: *
29: em_vi: *
30: em_htgo_hum: *
31: em_htgo_inv: *
32: em_htgo_rod: *
33: em_htg_hum: *
34: em_htg_inv: *
35: em_htg_rod: *
36: em_htg_other: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	15	62.5	17	6	AX099949
2	13.8	57.5	17	6	A31667
3	13.2	55.0	21	6	AX015842
4	13.2	55.0	23	6	AX146761
5	12.8	53.3	19	6	E05767
6	12.6	52.5	24	6	AR153077
7	12.4	51.7	21	6	AX095872
8	12.2	50.8	23	6	AX089242
9	12.2	50.8	20	6	AX183705
10	12.2	50.8	23	6	I20550
11	12.2	50.8	23	6	I20565
12	12.2	50.8	24	6	AX147428
13	12	50.0	20	6	AR092385
14	12	50.0	21	6	A63960
15	12	50.0	21	6	AR091470
16	11.8	49.2	17	6	AX029248
17	11.8	49.2	18	6	I78198
18	11.8	49.2	20	6	AX104199
19	11.8	49.2	24	6	AX116463
20	11.6	48.3	18	6	AR151248
21	11.6	48.3	20	6	A44370
22	11.6	48.3	20	6	AR103733
23	11.6	48.3	20	6	AR127296
24	11.6	48.3	20	6	I95598
25	11.6	48.3	24	6	AX022974
26	11.6	48.3	24	6	AX031224
27	11.4	47.5	21	6	AX117243
28	11.2	46.7	17	6	AR057574
29	11.2	46.7	17	6	AR115332
30	11.2	46.7	19	6	AR137256
31	11.2	46.7	19	6	AR156526
32	11.2	46.7	19	6	AX037378
33	11.2	46.7	20	6	A67865
34	11.2	46.7	20	6	AR051717
35	11.2	46.7	20	6	AR051719
36	11.2	46.7	20	6	AR067707
37	11.2	46.7	20	6	AR118894
38	11.2	46.7	20	6	I18250
39	11.2	46.7	20	6	I18252
40	11.2	46.7	21	6	AR016125
41	11.2	46.7	21	6	AR019123
42	11.2	46.7	21	12	AB068356
43	11.2	46.7	22	6	AR139330
44	11.2	46.7	22	6	AX118195
45	11.2	46.7	22	6	I17576

ALIGNMENTS

RESULT 1
AX099949/c 17 bp DNA
Sequence 9 from Patent WO0120034.
PAT 02-APR-2001
AX099949 GI:13538959

ORGANISM

house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE

Voss, J. and Timm, J.
Methods and compositions for the screening of cell cycle modulators
Patent: WO 0120034-A 9 22-MAR-2001;
BASF AKTRENSELSCHAFT (DE)
Location/Qualifiers

FEATURES

1..17
/organism="Mus musculus"
/db_xref="taxon:10090"

BASE COUNT 5 a 5 c 2 g 5 t
ORIGIN

Query Match 62.5%; Score 15; DB 6; Length 17;
Best Local Similarity 100.0%; Pred. No. 1.2e+04;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 tcaagcgtgaatga 19
|||||
Db 16 TCAAGCTGTAATGA 2

RESULT 2
LOCUS A31667 24 bp DNA PAT 11-DEC-1996
DEFINITION Synthetic pab D1.3 probe.
ACCESSION A31667
VERSION A31667.1 GI:1926461
KEYWORDS
SOURCE synthetic construct.
ORGANISM synthetic construct.
REFERENCE 1 (bases 1 to 24)
AUTHORS
TITLE METHODS FOR PRODUCING MEMBERS OF SPECIFIC BINDING PAIRS
JOURNAL Patent: WO 9201047-A 1 23-JAN-1992;
FEATURES
LOCATION/Qualifiers

BASE COUNT 6 a 7 c 3 g 8 t
ORIGIN
Query Match 57.5%; Score 13.8; DB 6; Length 24;
Best Local Similarity 88.2%; Pred. No. 4.6e+04;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 gtagcagcgtgtaat 17
|||||
Db 1 GTAGCAAGCCTATTAAT 17

RESULT 3
LOCUS AX015842 21 bp DNA PAT 07-SEP-2000
DEFINITION Sequence 41 from Patent WO9950285.
ACCESSION AX015842
VERSION AX015842.1 GI:10041602
KEYWORDS
SOURCE synthetic construct.
ORGANISM synthetic construct.
REFERENCE 1 (bases 1 to 21)
AUTHORS Griffiths,D.J., Weis,R.A. and Boyd,M.T.
TITLE Materials and methods relating to a novel retrovirus
JOURNAL Patent: WO 9950285-A 41 07-OCT-1999;
CANCER RES INST (GB); GRIFFITHS DAVID JOHN (GB); MATIILDA AND
TERENCE KENNEDY I (GB); VENABLES PATRICK JOHN WOODGATE (GB); WEISS
ROBERT ANTHONY (GB); BOYD MARK THOMAS (US)

FEATURES
SOURCE 1..21
Location/Qualifiers
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="Primer"
BASE COUNT 4 a 4 c 8 g 5 t
ORIGIN

Query Match 55.0%; Score 13.2; DB 6; Length 21;
Best Local Similarity 83.3%; Pred. No. 8.9e+04;
Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 gtagcagcgtgaatg 18
|||||
Db 4 GGAGTCAGGCTGTAATG 21

RESULT 4
LOCUS AX146761 23 bp DNA PAT 31-MAY-2001
DEFINITION Sequence 10 from Patent WO0134805.
ACCESSION AX146761
VERSION AX146761.1 GI:14285136
KEYWORDS
SOURCE synthetic construct.
ORGANISM synthetic construct.
REFERENCE 1 (bases 1 to 23)
AUTHORS Masters,D.N. and Vos,M.H.
TITLE Human vanilloid receptor gene
JOURNAL Patent: WO 0134805-A 10 17-MAY-2001;
Abbott Laboratories (US)

FEATURES
SOURCE 1..23
Location/Qualifiers
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="Primer"
BASE COUNT 8 a 4 c 6 g 5 t
ORIGIN

Query Match 55.0%; Score 13.2; DB 6; Length 23;
Best Local Similarity 83.3%; Pred. No. 8.9e+04;
Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 5 tcaagcgtgaatga 22
|||||
Db 18 TCAAGCTGCTTCATCA 1

RESULT 5
LOCUS E05767 19 bp DNA PAT 29-SEP-1997
DEFINITION Oligonucleotide to detect Campylobacter sp.
ACCESSION E05767
VERSION E05767.1 GI:2173954
KEYWORDS JP 1993276999-A/14.
SOURCE synthetic construct.
ORGANISM synthetic construct.
REFERENCE 1 (bases 1 to 19)
AUTHORS Matsunaka,A., Kagawa,S., Otsuka,N., Yamashita,K., Kumagai,S. and
Satou,M.

TITLE OLIGONUCLEOTIDE FOR DETECTING BACTERIUM BELONGING TO GENUS
CAMPYLOBACTER, METHOD FOR DETECTING BACTERIUM BELONGING TO GENUS
CAMPYLOBACTER AND REAGENT KIT FOR DETECTION
JOURNAL Patent: JP 1993276999-A 14 26-OCT-1993;
TOYOBO CO LTD
COMMENT

OS Artificial gene
OC Artificial sequence; Genes.
PN JP 1993276999-A/14
PD 26-OCT-1993
PF 02-APR-1992 JP 1992080769
PI MATSUOKA AKIRA, KAGAWA SHOHEI, OTSUKA NORIMITSU, PI
YAMASHITA KEIKO,
PC KUMAGAI SHYUOKO, SATOU MIYUKI
CC C12Q1/68,C07H21/04,C12N15/10,C12Q1/04,(C12Q1/04,C12R1.01); CC
strandedness: Single;
CC topology: Linear;
CC Feature is identified by similarity;
FH key Location/Qualifiers
FT misc-feature 1..19
FT /note="this sequence is complementary to FT

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 5, 2002, 19:16:13 ; Search time 1481.91 Seconds
(without alignments)
174.031 Million cell updates/sec

Title: US-09-424-705-7
Perfect score: 24
Sequence: 1 gtagcaagcgtgataatgcatc 24

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues
Total number of hits satisfying chosen parameters: 13210

Minimum DB seq length: 0
Maximum DB seq length: 24

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: em_estfun:*
2: em_esthum:*
3: em_estin:*
4: em_estom:*
5: em_estpl:*
6: em_estda:*
7: em_estro:*
8: em_estov:*
9: em_hic:*
10: gb_est1:*
11: gb_est2:*
12: gb_hic:*
13: gb_gss:*
14: em_gss_fun:*
15: em_gss_hum:*
16: em_gss_inv:*
17: em_gss_pln:*
18: em_gss_pro:*
19: em_gss_rod:*
20: em_gss_vrt:*
21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	11	45.8	22	13	AZ833359 2M0115G19
C 2	10.8	45.0	24	13	TA211D02P
C 3	10.4	43.3	19	13	AZ507472
C 4	10.4	43.3	19	13	AZ772566
C 5	10.4	43.3	20	13	AZ827586 2M0104C08
C 6	10.2	42.5	24	13	AZ416129 2M0191D16
C 7	10.2	41.7	22	13	AZ778745 2M0014M08
C 8	9.8	40.8	19	13	AZ623412 1M0461A08
C 9	9.8	40.8	20	13	AZ328275 1M0052A01
C 10	9.8	40.8	21	13	AZ825240 2M0100N16
C 11	9.8	40.8	22	13	AZ400303 1M0166H05
C 12	9.8	40.8	22	13	AZ416988 1M0192P23

RESULT	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	
AZ833359/c	9.8	40.8	23	13	AZ445661	9.8	40.8	23	13	TA184F06Q	9.8	40.8	23	13	TA184F06Q	9.8	40.8	23	13	TA184F06Q	9.8	40.8	23	13	TA184F06Q
LOCUS	9.8	40.8	23	13	TA184F06Q	9.8	40.8	23	13	TA184F06Q	9.8	40.8	23	13	TA184F06Q	9.8	40.8	23	13	TA184F06Q	9.8	40.8	23	13	TA184F06Q
DEFINITION	9.8	40.8	23	13	TA184F06Q	9.8	40.8	23	13	TA184F06Q	9.8	40.8	23	13	TA184F06Q	9.8	40.8	23	13	TA184F06Q	9.8	40.8	23	13	TA184F06Q
ACCESSION	9.8	40.8	23	13	TA184F06Q	9.8	40.8	23	13	TA184F06Q	9.8	40.8	23	13	TA184F06Q	9.8	40.8	23	13	TA184F06Q	9.8	40.8	23	13	TA184F06Q
VERSION	9.8	40.8	23	13	TA184F06Q	9.8	40.8	23	13	TA184F06Q	9.8	40.8	23	13	TA184F06Q	9.8	40.8	23	13	TA184F06Q	9.8	40.8	23	13	TA184F06Q
KEYWORDS	9.8	40.8	23	13	TA184F06Q	9.8	40.8	23	13	TA184F06Q	9.8	40.8	23	13	TA184F06Q	9.8	40.8	23	13	TA184F06Q	9.8	40.8	23	13	TA184F06Q
SOURCE	9.8	40.8	23	13	TA184F06Q	9.8	40.8	23	13	TA184F06Q	9.8	40.8	23	13	TA184F06Q	9.8	40.8	23	13	TA184F06Q	9.8	40.8	23	13	TA184F06Q
ORGANISM	9.8	40.8	23	13	TA184F06Q	9.8	40.8	23	13	TA184F06Q	9.8	40.8	23	13	TA184F06Q	9.8	40.8	23	13	TA184F06Q	9.8	40.8	23	13	TA184F06Q
REFERENCE	9.8	40.8	23	13	TA184F06Q	9.8	40.8	23	13	TA184F06Q	9.8	40.8	23	13	TA184F06Q	9.8	40.8	23	13	TA184F06Q	9.8	40.8	23	13	TA184F06Q
AUTHORS	9.8	40.8	23	13	TA184F06Q	9.8	40.8	23	13	TA184F06Q	9.8	40.8	23	13	TA184F06Q	9.8	40.8	23	13	TA184F06Q	9.8	40.8	23	13	TA184F06Q
TITLE	9.8	40.8	23	13	TA184F06Q	9.8	40.8	23	13	TA184F06Q	9.8	40.8	23	13	TA184F06Q	9.8	40.8	23	13	TA184F06Q	9.8	40.8	23	13	TA184F06Q
JOURNAL	9.8	40.8	23	13	TA184F06Q	9.8	40.8	23	13	TA184F06Q	9.8	40.8	23	13	TA184F06Q	9.8	40.8	23	13	TA184F06Q	9.8	40.8	23	13	TA184F06Q
COMMENT	9.8	40.8	23	13	TA184F06Q	9.8	40.8	23	13	TA184F06Q	9.8	40.8	23	13	TA184F06Q	9.8	40.8	23	13	TA184F06Q	9.8	40.8	23	13	TA184F06Q

ALIGNMENTS

RESULT 1
AZ833359 22 bp DNA GSS 20-FEB-2001
2M0115G19F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
clone UUGC2M0115G19 F, DNA sequence.

ACCESSION AZ833359
VERSION AZ833359.1 GI:13003267
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Rellly,
M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
and Wright,D., Weiss,R.
Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
Unpublished (2000)
Contact: Robert B. Weiss
University of Utah Genome Center
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert length: 10000 Std Error: 0.00
Plate: 0115 row: G column: 19
Seq primer: CCTGTAAACGACGCCACT
Class: Plasmid ends
High quality sequence stop: 22.
Location/Qualifiers
1. .22

```

/organism="Mus musculus"
/strain="C57BL/6J"
/db.xref="taxon:10090"
/clone="UUCGM0115G19"
/clone_1lb="Mouse 10kb plasmid UUCGM library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/notes="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adapted DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD42 (g14732114|gplAF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

```

```

BASE COUNT      8 a      8 c      1 g      5 t
ORIGIN

```

```

Query Match      45.8%; Score 11; DB 13; Length 22;
Best Local Similarity 100.0%; Pred. No. 3.8e+05;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 9 ggcgtgaatga 19
    |||||
Db 20 GCGTGAATGA 10

```

```

RESULT 2
TA211D02P/c 24 bp DNA GSS 13-DEC-2000
LOCUS T. brucei sheared genomic DNA clone 211d02, forward sequence,
DEFINITION genomic survey sequence.
ACCESSION AL478562
VERSION AL478562.1 GI:11644521
KEYWORDS GSS.
SOURCE Trypanosoma brucei.
ORGANISM Trypanosoma brucei.
Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
Trypanosoma.
1 (bases 1 to 24)

```

```

REFERENCE Hall,N., Bowman,S., Lennard,N.J., Doggett,J., Atkin,R.,
AUTHORS Chillingworth,C., Ormond,D., Harris,B., El-Sayed,N., Hou,L.,
Melville,S.E., Rajandream,M.A. and Barrell,B.G.

```

```

TITLE Direct Submission
JOURNAL Submitted (10-DEC-2000) Trypanosoma brucei genome sequencing
project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton,
Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk and
nilesanger.ac.uk

```

```

COMMENT Constructed at the Institute for Genomic Research (TIGR),
Rockville, MD. Genomic DNA isolated from a cloned population of
Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically sheared
to give a tight size distribution (
4 kb). The v+1 method used for the library construction is
described in detail in Smith, H. and Venter, J.C. (Making small
insert libraries for whole genome shotgun sequencing projects. In
Genome Sequencing: A Practical Approach, eds. M. Vaubin and B.
Barrell, Oxford University Press, 1999).
Email: nilesayed@tigr.org
Details of T. brucei sequencing at the Sanger Centre are available
at http://www.sanger.ac.uk/Projects/T_brucei/.
Location/Qualifiers

```

FEATURES.

```

source
1. .24
/organism="Trypanosoma brucei"
/strain="TREU927"
/db.xref="taxon:5691"
/clone="211d02"
BASE COUNT      3 a      8 c      5 g      8 t
ORIGIN

```

```

Query Match      45.0%; Score 10.8; DB 13; Length 24;
Best Local Similarity 68.2%; Pred. No. 4.7e+05;
Matches 15; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

```

```

Qy 3 agtaagcgtgaatgacac 24
    ||| ||| ||| ||| |||
Db 24 AGTAAGGCGAGTATAGACCACC 3

```

```

RESULT 3
A2507472/c 19 bp DNA GSS 05-OCT-2000
LOCUS IM0349D11F Mouse 10kb plasmid UUCGM library Mus musculus genomic
DEFINITION clone UUCGM0349D11 F, DNA sequence.
ACCESSION A2507472
VERSION A2507472.1 GI:10688788
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 19)
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamill,C.,
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M.,
Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
and Wright,D., Weiss,R.
Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
Unpublished (2000)
Contact: Robert B. Weiss
University of Utah Genome Center
Km. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0349 row: D column: 11
Seq primer: CGTTGTAAACGACGGCCAGT
Class: plasmid ends
High quality sequence stop: 19.
Location/Qualifiers

```

```

TITLE Mouse whole genome scaffolding with paired end reads from 10kb
JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
Km. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0349 row: D column: 11
Seq primer: CGTTGTAAACGACGGCCAGT
Class: plasmid ends
High quality sequence stop: 19.
Location/Qualifiers

```

```

FEATURES
source
1. .19
/organism="Mus musculus"
/strain="C57BL/6J"
/db.xref="taxon:10090"
/clone="UUCGM0349D11"
/clone_1lb="Mouse 10kb plasmid UUCGM library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/notes="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adapted DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD42 (g14732114|gplAF129072.1), a copy-number

```

```

of pMD42 (g14732114|gplAF129072.1), a copy-number

```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 5, 2002, 20:07:18 ; Search time 168.74 Seconds
(without alignments)
121.938 Million cell updates/sec

Title: US-09-424-705-7

Sequence: 1 gtagtcaagctgtaatgcatc 24

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

number of hits satisfying chosen parameters: 607738

Minimum DB seq length: 0
Maximum DB seq length: 24

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1: /SIDSL/gcgdata/geneseq/geneseq/NA1980.DAT:*
2: /SIDSL/gcgdata/geneseq/geneseq/NA1981.DAT:*
3: /SIDSL/gcgdata/geneseq/geneseq/NA1982.DAT:*
4: /SIDSL/gcgdata/geneseq/geneseq/NA1983.DAT:*
5: /SIDSL/gcgdata/geneseq/geneseq/NA1984.DAT:*
6: /SIDSL/gcgdata/geneseq/geneseq/NA1985.DAT:*
7: /SIDSL/gcgdata/geneseq/geneseq/NA1986.DAT:*
8: /SIDSL/gcgdata/geneseq/geneseq/NA1987.DAT:*
9: /SIDSL/gcgdata/geneseq/geneseq/NA1988.DAT:*
10: /SIDSL/gcgdata/geneseq/geneseq/NA1989.DAT:*
11: /SIDSL/gcgdata/geneseq/geneseq/NA1990.DAT:*
12: /SIDSL/gcgdata/geneseq/geneseq/NA1991.DAT:*
13: /SIDSL/gcgdata/geneseq/geneseq/NA1992.DAT:*
14: /SIDSL/gcgdata/geneseq/geneseq/NA1993.DAT:*
15: /SIDSL/gcgdata/geneseq/geneseq/NA1994.DAT:*
16: /SIDSL/gcgdata/geneseq/geneseq/NA1995.DAT:*
17: /SIDSL/gcgdata/geneseq/geneseq/NA1996.DAT:*
18: /SIDSL/gcgdata/geneseq/geneseq/NA1997.DAT:*
19: /SIDSL/gcgdata/geneseq/geneseq/NA1998.DAT:*
20: /SIDSL/gcgdata/geneseq/geneseq/NA1999.DAT:*
21: /SIDSL/gcgdata/geneseq/geneseq/NA2000.DAT:*
22: /SIDSL/gcgdata/geneseq/geneseq/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	24	100.0	24	19	AAV73336	Mouse OKT3 variant
2	15	62.5	17	22	AAV57365	Murine Cd25A intr
3	13.8	57.5	24	13	AAO23468	Probe DL3CDR3A fo
4	13.6	56.7	24	15	AAO81088	Mutagenic primer t
5	13.2	55.0	21	20	AAZ32535	Human retrovirus-5
6	13.2	55.0	22	19	AAV49973	PCR primer for hum
7	13.2	55.0	22	21	AAV82713	Human IGA nephropa
8	13.2	55.0	23	22	AAO58830	Human vanilloid re
9	12.8	53.3	19	14	AAO61580	Campylobacter bact
10	12.8	53.3	21	22	AAV6285	Human gene single
11	12.8	53.3	24	19	AAV49937	PCR primer for hum

12	12.8	53.3	24	21	AAV82677	Human IGA nephropa
13	12.6	52.5	21	18	AAV45049	Helicobacter CP2 a
14	12.6	52.5	24	18	AAV00538	PCR primer V-alpha
15	12.6	52.5	24	20	AAZ25493	Primer dependent d
16	12.6	52.5	24	21	AAV86868	Upstream probe for
17	12.4	51.7	24	17	AAV17774	Antisense primer f
18	12.2	50.8	19	22	AAV83994	Rat hedgehog prote
19	12.2	50.8	20	19	AAV18283	Measles virus L pr
20	12.2	50.8	20	20	AAZ22917	Primer specific fo
21	12.2	50.8	20	22	AAV92020	B thuringiensis 45
22	12.2	50.8	23	14	AAO51140	C. tropicalis 4443
23	12.2	50.8	23	22	AAV91097	Human inflammatory
24	12.2	50.8	24	22	AAH42174	Human Jagged2 (Ser
25	12	50.0	20	21	AAZ38521	Human microtubule-
26	11.8	49.2	17	19	AAV64145	Rat activin recept
27	11.8	49.2	18	16	AAV57674	RSV N hairpin ribo
28	11.8	49.2	18	22	AAV63025	Shrimp white spot
29	11.8	49.2	18	22	AAV60145	Human ATM gene exo
30	11.8	49.2	20	20	AAZ02589	PCR primer used to
31	11.8	49.2	20	20	AAV38474	E. coli SecA antis
32	11.8	49.2	20	21	AAZ48442	C. jejuni nucleic
33	11.8	49.2	20	22	AAV99262	Immunostimulatory
34	11.8	49.2	21	21	AAV58066	Porphyromonas ging
35	11.8	49.2	21	21	AAV58070	Porphyromonas ging
36	11.8	49.2	23	14	AAO51139	C. tropicalis 4414
37	11.8	49.2	23	21	AAV15611	Reverse PCR primer
38	11.8	49.2	24	21	AAV57416	Human cell cycle r
39	11.8	49.2	24	21	AAV57417	Human cell cycle r
40	11.6	48.3	18	14	AAV38790	SNP specific lower
41	11.6	48.3	18	14	AAV49332	Degenerate oligonu
42	11.6	48.3	18	18	AAV93475	DOAI allele determ
43	11.6	48.3	18	18	AAV93476	DOAI allele determ
44	11.6	48.3	18	20	AAV90255	DOAI gene PCR prim
45	11.6	48.3	18	20	AAV90254	DOAI gene PCR prim

ALIGNMENTS

RESULT 1	
AAV73336	standard; DNA: 24 BP.
ID	AAV73336
XX	
AC	AAV73336:
XX	
DT	26-FEB-1999 (first entry)
XX	
DE	Mouse OKT3 variant antibody primer SK1 DNA.
XX	
KM	OKT3: monoclonal antibody; Mab: point mutation; transplant rejection;
KW	organ recipient; diagnosis: tumour; therapy: primer; ss.
XX	
OS	Synthetic.
OS	Mus sp.
XX	
PN	DEL9721700-CL.
XX	
PD	19-NOV-1998.
XX	
PF	23-MAY-1997: 97DE-1021700.
XX	
PR	23-MAY-1997: 97DE-1021700.
XX	
PA	(DEKR-) DEUT KREBSFORSCHUNGSZENTRUM.
XX	
PI	Kipriyanov S, Little M, Moldenhauer G;
XX	
DR	WPI: 1998-596150/51.
XX	
PT	Monoclonal antibody OKT3 with point-mutation - where cysteine is
PT	replaced by another polar amino acid, useful for controlling
PT	transplant rejection, and in tumour diagnostics and therapy
XX	

PS Claim 7; Column 3; 8pp; German.

XX This sequence is a primer used in the construction of a monoclonal
CC antibody (MAB) derived from OKT3 with a point-mutation where Cys at
CC position H100A is replaced with another polar amino acid, in this example
CC Ser. The MAB is used in lowering or eliminating the transplant rejection
CC in an organ recipient and for diagnostic methods for tumours and tumour
CC therapy.

XX Sequence 24 BP; 7 A; 4 C; 6 G; 7 T; 0 other;

Query Match

Best Local Similarity 100.0%; Score 24; DB 19; Length 24;
Pred. No. 0.019;

Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gtagtcaagcgtgtatgcatc 24
|||||
Db 1 gtagtcaagcgtgtatgcatc 24

RESULT 2

AAE57365/C
ID AAE57365 standard; DNA; 17 BP.

AC AAE57365;

DT 11-JUN-2001 (first entry)

DE Murine Cdc25A intron 3/exon 4 splice junction sequence.

KW Cdc25; Cdc25 phosphatase; transcription; modulator; murine; CDC25A;
KW exon; intron; ds.

OS Mus sp.

PN WO200120034-A2.

PD 22-MAR-2001.

PE 11-SEP-2000; 2000WO-US24838.

PR 13-SEP-1999; 99US-0153639.

PA (BADI) BASF AG.

PI Voss J, Timm J;

WP; 2001-244825/25.

XX Assay for screening modulators of Cdc25 activity by using a cell having
PT a recombinant Cdc25 phosphatase gene whose expression alters the
PT transcription of a selected gene in the presence of a modulator -

XX Example 1; Page 15; 55pp; English.

XX The invention relates to a method of identifying a modulator of Cdc25
CC activity that comprises contacting a test cell having a recombinant Cdc25
CC phosphatase gene whose expression alters transcription of a selected
CC gene, with a compound under conditions where recombinant Cdc25
CC phosphatase gene is expressed and alters the transcription of a selected
CC gene as an indication of the compound being a modulator of Cdc25-mediated
CC transcription. The method is useful for identifying modulators of Cdc25
CC activity. Sequences AAE57363-376 represent intron/exon splice junction
CC sequences of the murine Cdc25A gene.

XX Sequence 17 BP; 5 A; 5 C; 2 G; 5 T; 0 other;

Query Match

Best Local Similarity 62.5%; Score 15; DB 22; Length 17;
Pred. No. 3.5e+02;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 tcaagcgtgtatga 19
|||||
Db 16 TCAGGCTGTATGA 2

RESULT 3

AAQ23468
ID AAQ23468 standard; DNA; 24 BP.

AC AAQ23468;

DT 21-MAY-1992 (first entry)

DE Probe D1.3CDR3A for pAb D1.3.

KW Fd; bacteriophage; gene III; filamentous; phagemid; capsid; coat;
KW plus; g3p; binding; adsorption; gene VIII; diverse repertoire;
KW specific binding pairs; replicable genetic display package;
KW immunoglobulin; ss.

OS Synthetic.

PN WO9201047-A.

PD 23-JAN-1992.

PE 10-JUL-1991; 91WO-GB01134.

PR 15-MAY-1991; 91GB-0010549.

PR 10-JUL-1990; 90GB-0015198.

PR 19-OCT-1990; 90GB-0022845.

PR 12-NOV-1990; 90GB-0024503.

PR 06-MAR-1991; 91GB-0004744.

PA (CAMP-) CAMBRIDGE ANTIBODY.

PA (MEDI-) MED RES COUNCIL.

PI McCafferty J, Pope AR, Johnson KS, Hoogenboom HRJ, Griffiths AD;

PI Jackson RH, Holliger KP, Marks JD, Clackson TP, Chiswell DJ;

PI Winter GP, Bonnett TP;

DR WP; 1992-056862/07.

XX Producing members of specific binding pairs - by expression in
PT recombinant host cells with a secreting replicable genetic
PT display package.

XX Example 10; Page 57; 209pp; English.

XX The probe was used to distinguish between pAb D1.3 and pAb NQ11
CC which express antibody fragments specific for hen egg lysozyme and
CC oxazalone respectively. The probe was used to monitor the
CC enrichment of pAb D1.3 from a mixed vector population contg. pAb
CC NQ11 by application to a lysozyme sepharose column. An enrichment
CC of 1000 fold was achieved in one round.

CC See also AAQ21092-100, 103-116, 126-131; AAQ23463, 465-495, 693-719,
CC 736-738, and 793-863.

XX Sequence 24 BP; 6 A; 7 C; 3 G; 8 T; 0 other;

XX Query Match 57.5%; Score 13.8; DB 13; Length 24;
Best Local Similarity 88.2%; Pred. No. 1.4e+03;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 gtagtcaagcgtgtat 17
|||||
Db 1 gtagtcaagcgtgtat 17

RESULT 4

AAQ81088/C
ID AAQ81088 standard; DNA; 24 BP.

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 5, 2002, 19:19:13 ; Search time 66.33 Seconds
(without alignments)
81.946 Million cell updates/sec

Title: US-09-424-705-7

Perfect score: 24

Sequence: 1 gtacgaagctgtatgatcatc 24

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 351203 seqs, 113238999 residues

T number of hits satisfying chosen parameters: 281400

Minimum DB seq length: 0
Maximum DB seq length: 24

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A.COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B.COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A.COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B.COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS.COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	12.6	52.5	24	US-09-358-972-78	Sequence 78, Appl
C 2	12.4	51.7	24	PCT-US95-10973A-64	Sequence 64, Appl
C 3	12.2	50.8	23	US-07-963-290A-5	Sequence 5, Appl
C 4	12.2	50.8	23	US-07-963-290A-24	Sequence 24, Appl
C 5	12.2	50.8	23	US-08-819-867-38	Sequence 38, Appl
C 6	12	50.0	20	US-09-289-368-56	Sequence 56, Appl
C 7	12	50.0	21	US-08-781-620B-4	Sequence 4, Appl
C 8	11.8	49.2	18	US-08-334-847-905	Sequence 905, Appl
C 9	11.8	49.2	23	US-08-819-867-37	Sequence 37, Appl
C 10	11.8	49.2	23	US-08-766-354A-21	Sequence 21, Appl
C 11	11.6	48.3	18	US-07-861-458C-74	Sequence 74, Appl
C 12	11.6	48.3	20	US-08-332-747-26	Sequence 26, Appl
C 13	11.6	48.3	20	US-09-009-913-257	Sequence 257, Appl
C 14	11.6	48.3	20	US-09-429-093-2	Sequence 2, Appl
C 15	11.6	48.3	24	US-09-230-804-24	Sequence 24, Appl
C 16	11.4	47.5	17	US-08-753-147-92	Sequence 92, Appl
C 17	11.2	46.7	17	US-08-292-620A-1778	Sequence 1778, Appl
C 18	11.2	46.7	17	US-08-985-162-187	Sequence 187, Appl
C 19	11.2	46.7	17	US-09-071-845-1778	Sequence 1778, Appl
C 20	11.2	46.7	19	US-09-050-159-3	Sequence 3, Appl
C 21	11.2	46.7	19	US-09-407-818-10	Sequence 10, Appl
C 22	11.2	46.7	20	US-08-343-785-34	Sequence 34, Appl
C 23	11.2	46.7	20	US-08-343-785-36	Sequence 36, Appl
C 24	11.2	46.7	20	US-08-462-321-34	Sequence 34, Appl
C 25	11.2	46.7	20	US-08-462-321-36	Sequence 36, Appl
C 26	11.2	46.7	20	US-08-447-031A-3	Sequence 3, Appl
C 27	11.2	46.7	20	US-08-946-458-34	Sequence 34, Appl

28	11.2	46.7	20	3	US-08-946-458-36	Sequence 36, Appl
29	11.2	46.7	20	3	US-08-765-340-20	Sequence 20, Appl
C 30	11.2	46.7	20	4	US-09-489-868A-27	Sequence 27, Appl
C 31	11.2	46.7	21	1	US-08-531-556-13	Sequence 13, Appl
C 32	11.2	46.7	21	1	US-08-472-416-13	Sequence 13, Appl
C 33	11.2	46.7	22	1	US-08-062-368-9	Sequence 9, Appl
C 34	11.2	46.7	22	2	US-08-474-450A-7	Sequence 7, Appl
C 35	11.2	46.7	22	4	US-08-983-466-38	Sequence 38, Appl
C 36	11.2	46.7	23	4	US-09-092-063-2	Sequence 2, Appl
C 37	11.2	46.7	24	4	US-07-741-940-74	Sequence 74, Appl
C 38	11.2	45.8	20	1	US-07-991-867B-50	Sequence 50, Appl
C 39	11.2	45.8	20	1	US-08-289-348A-74	Sequence 74, Appl
C 40	11.2	45.8	20	1	US-08-452-654-74	Sequence 74, Appl
C 41	11.2	45.8	20	1	US-08-452-655B-74	Sequence 74, Appl
C 42	11.2	45.8	20	1	US-08-452-655B-74	Sequence 74, Appl
C 43	11.2	45.8	20	2	US-08-544-332-50	Sequence 50, Appl
C 44	11.2	45.8	20	3	US-08-450-582-74	Sequence 74, Appl
C 45	11.2	45.8	21	4	US-09-487-368A-11	Sequence 11, Appl

ALIGNMENTS

RESULT 1
US-09-358-972-78/c
Sequence 78, Application US/09358972
Patent No. 6235480
GENERAL INFORMATION:
APPLICANT: Shultz, John W.
APPLICANT: Lewis, Martin K.
APPLICANT: Lieppe, Donna
APPLICANT: Mandrekar, Michelle
APPLICANT: Kephart, Daniel
APPLICANT: Rhodes, Richard B.
APPLICANT: Andrews, Christine A.
APPLICANT: Hartnett, James R.
APPLICANT: Gu, Trent
APPLICANT: Olson, Ryan J.
APPLICANT: Wood, Keith W.
APPLICANT: Welch, Roy
TITLE OF INVENTION: Nucleic Acid Detection
FILE REFERENCE: Pro-103 6868/75528
CURRENT APPLICATION NUMBER: US/09/358, 972
CURRENT FILING DATE: 1999-07-22
EARLIER APPLICATION NUMBER: 09/252, 436
EARLIER FILING DATE: 1999-02-18
EARLIER APPLICATION NUMBER: 09/042, 287
EARLIER FILING DATE: 1998-03-13
NUMBER OF SEQ ID NOS: 290
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 78
LENGTH: 24
TYPE: DNA
ORGANISM: Upstream probe for kanamycin
US-09-358-972-78

Query Match 52.5%; Score 12.6; DB 4; Length 24;
Best Local Similarity 78.9%; Pred. No. 8.4e+02;
Matches 15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Db 19 GTCAAGACTTACGATCA 1

Qy 4 gtcaagctgtatgatca 22
||||| || |||||

RESULT 2
PCT-US95-10973A-64/c
Sequence 64, Application PC/TUS9510973A
GENERAL INFORMATION:
APPLICANT: Pitzm Pharmaceuticals, Inc.
TITLE OF INVENTION: CONJUGATES OF VASCULAR ENDOTHELIAL GROWTH FACTOR WITH TARGE
NUMBER OF SEQUENCES: 107

CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/10973A
FILING DATE: 29-AUG-1995
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Nottenburg, Carol
REGISTRATION NUMBER: 39,317
REFERENCE/DOCKET NUMBER: 760100.413PC
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 64:
SEQUENCE CHARACTERISTICS:
LENGTH: 24 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
PCT-US95-10973A-64

Query Match 51.7%; Score 12.4; DB 5; Length 24;
Best Local Similarity 72.7%; Pred. No. 1.1e+03;
Matches 16; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
Qy 3 agtcaagctgaatgcatc 24
||| ||||| ||||| |||
Db 23 AGCCGAGCGGTGATGAGATC 2

RESULT 3
US-07-963-290A-5
Sequence 5, Application US/07963290A
Patent No. 5514586
GENERAL INFORMATION:
APPLICANT: HOTTINGER, Herbert
APPLICANT: NIEDERBERGER, Peter
APPLICANT: PRIMORE, David
APPLICANT: STAEGER-ROOS, Ursula
TITLE OF INVENTION: FOOD ADDITIVES
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESSEE: VOGT & O'DONNELL
STREET: 707 WESTCHESTER AVENUE
CITY: WHITE PLAINS
STATE: NEW YORK
COUNTRY: USA
ZIP: 10604
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/963,290A
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 8923998
FILING DATE: 25-OCT-1989
ATTORNEY/AGENT INFORMATION:

NAME: HARACZ, Stephen M
REGISTRATION NUMBER: 33,397
REFERENCE/DOCKET NUMBER: NO 2739/39/01
TELECOMMUNICATION INFORMATION:
TELEPHONE: 914 328-0055
TELEFAX: 914 328-0060
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 23 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "View of the 180bp TaqI-SphI
DESCRIPTION: fragment from plasmid Ecl carrying a part of the EMV3 gene"
FEATURE:
NAME/KEY: -
LOCATION: 9..11
OTHER INFORMATION: /note= "TaqI restriction site used
OTHER INFORMATION: to digest plasmid Ecl. The SphI restriction site is not sh
US-07-963-290A-5

Query Match 50.8%; Score 12.2; DB 1; Length 23;
Best Local Similarity 82.4%; Pred. No. 1.3e+03;
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Qy 8 agtctgaatgcatc 24
| ||||| ||||| |||
Db 1 ATGCTGTAATGAAATATC 17

RESULT 4
US-07-963-290A-24/C
Sequence 24, Application US/07963290A
Patent No. 5514586
GENERAL INFORMATION:
APPLICANT: HOTTINGER, Herbert
APPLICANT: NIEDERBERGER, Peter
APPLICANT: PRIMORE, David
APPLICANT: STAEGER-ROOS, Ursula
TITLE OF INVENTION: FOOD ADDITIVES
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESSEE: VOGT & O'DONNELL
STREET: 707 WESTCHESTER AVENUE
CITY: WHITE PLAINS
STATE: NEW YORK
COUNTRY: USA
ZIP: 10604
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/963,290A
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 8923998
FILING DATE: 25-OCT-1989
ATTORNEY/AGENT INFORMATION:
NAME: HARACZ, Stephen M
REGISTRATION NUMBER: 33,397
REFERENCE/DOCKET NUMBER: NO 2739/39/01
TELECOMMUNICATION INFORMATION:
TELEPHONE: 914 328-0055
TELEFAX: 914 328-0060
INFORMATION FOR SEQ ID NO: 24:
SEQUENCE CHARACTERISTICS:
LENGTH: 23 base pairs
TYPE: nucleic acid